

Validación de criterio de la metodología de empleo del análisis estadístico implicativo en los estudios de causalidad en salud

Criteria Validation of the Methodology for the Use of Implicative Statistical Analysis in Health Causality Studies

Dra. C. Nelsa María Sagaró del Campo^{1*}

0000-0002-1964-8830

Dra. C. Larisa Zamora Matamoros²

0000-0003-2210-0806

¹ Universidad de Ciencias Médicas de Santiago de Cuba, Santiago de Cuba, Cuba.

² Universidad de Oriente, Santiago de Cuba, Cuba.

* Autor para la correspondencia: nelsa@infomed.sld.cu

RESUMEN

Introducción: el diseño de la metodología ASI-IMC permite una correcta aplicación del análisis estadístico implicativo en los estudios de causalidad en salud. Luego surgió la necesidad de validarla.

Objetivo: evaluar la efectividad de la metodología de aplicación del ASI-IMC.

Métodos: se realizó un estudio observacional analítico prospectivo de tipo casos y controles anidado en una cohorte, cuyo universo de estudio quedó conformado por todas las mujeres mayores de 18 años de edad, con el diagnóstico clínico e histológico de cáncer de mama, procedentes de la provincia de Santiago de Cuba, atendidas en el Hospital Oncológico “Conrado Benítez”, entre 2014 y 2019. Se emplearon como covariables 25 factores pronósticos supuestos. Se aplicó la regresión logística binaria previa verificación del cumplimiento de los supuestos requeridos sobre una muestra de 280 pacientes a razón de un control por caso que constituyó el mismo conjunto de datos al que se aplicó el análisis estadístico implicativo, para luego comparar los resultados de ambas técnicas. Se consideró la regresión como el estándar de oro, para lo cual se estimaron 14 indicadores: sensibilidad, especificidad, valores predictivos, razones de verosimilitud, odds ratio de diagnóstico, entre otros.

Resultados: fueron identificados, por ambas técnicas estadísticas, como factores de buen pronóstico de mortalidad por cáncer de mama en la población estudiada la determinación



de los biomarcadores y de mal pronóstico el estadio avanzado, la metástasis y la quimioterapia. Los indicadores de eficacia arrojaron valores a favor de la técnica evaluada.

Conclusiones: se validó de manera satisfactoria la metodología diseñada demostrando ser efectiva para la identificación de factores pronósticos.

Palabras clave: análisis estadístico implicativo; factores pronósticos; cáncer de mama; regresión logística.

ABSTRACT

Introduction: the design of the ASI-IMC methodology allows a correct application of the implicative statistical analysis in the studies of causality in health. Then the need arose to validate it.

Objective: to validate the designed ASI-IMC methodology.

Methods: a prospective analytical observational study of the case and control type nested in a cohort was carried out. The universe of study was made up of all women over 18 years of age, with a clinical and histological diagnosis of breast cancer, from the province of Santiago de Cuba, treated at the "Conrado Benítez" Oncological Hospital, between 2014 and 2019. Twenty-five assumed prognostic factors were used as covariates. Binary logistic regression was applied after verification of compliance with the required assumptions on a sample of 280 patients at the rate of one control per case, which constituted the same data set to which the statistical analysis was applied, in order to compare the results of both techniques. Regression was considered as the gold standard, for which 14 indicators were estimated: sensitivity, specificity, predictive values, likelihood ratios, diagnostic odds ratio, among others.

Results: both statistical techniques identified biomarkers as good prognosis factors for breast cancer mortality in the study population, and advanced stage, metastasis, and chemotherapy as poor prognostic factors. The efficacy indicators showed values in favor of the evaluated technique.

Conclusions: the designed methodology was satisfactorily validated, proving to be effective for the identification of prognostic factors.

Keywords: implicative statistical analysis; prognostic factors; breast cancer; logistic regression.

Recibido: 08/07/2021

Aprobado: 17/03/2022

Introducción

El análisis estadístico implicativo surgió en los años 80 para resolver problemas de la Didáctica de las matemáticas.^(1,2) Según se expresa en Sagaró y Zamora,³ tras el



reconocimiento de las posibilidades de empleo de esta técnica en las investigaciones médicas de causalidad, se propuso a partir de la introducción de nuevos procedimientos una metodología denominada ASI-IMC (análisis estadístico implicativo en las investigaciones médicas de causalidad) y son varias las investigaciones que la han empleado para identificar factores de riesgo y pronósticos.⁴

Una de las formas de validación de la metodología ASI-IMC fue la comparación con un criterio externo definido como estándar, la cual se presenta en este trabajo. El estándar de oro empleado fue la regresión logística binaria por ser la técnica más empleada en estos estudios desde mediados del siglo XX.

Se escogió para exponer este tipo de validación un estudio de identificación de factores pronósticos de mortalidad por cáncer de mama, ya que esta enfermedad constituye un problema de salud prioritario en el país y en particular en la provincia de Santiago de Cuba, al estar entre las primeras causas de muerte de manera sostenida y con tendencia ascendente desde el 2008, todo lo cual es corroborado por la literatura y otros estudios realizados en la provincia.⁵

El objetivo de este trabajo es evaluar la efectividad de la metodología de aplicación del ASI y, colateralmente, identificar los factores pronósticos de mortalidad por cáncer de mama en el territorio estudiado.

Métodos

Para evaluar la efectividad de la metodología ASI-IMC se realizó un estudio observacional analítico prospectivo de tipo casos y controles anidado en una cohorte cuyo universo de estudio quedó conformado por todas las mujeres mayores de 18 años de edad con el diagnóstico clínico e histológico de cáncer de mama, procedentes de la provincia de Santiago de Cuba, que fueron atendidas en el Hospital Oncológico “Conrado Benítez”, en el período de junio de 2014 a abril de 2019, constituyendo los casos todas las fallecidas al final del período de estudio y los controles una muestra de las vivas en igual momento.

Se tomó un control por cada caso para una razón 1:1, por lo que se empleó la fórmula de tamaño muestral para casos y controles balanceados que se calculó a partir de los resultados de los estudios pilotos realizados.^(6,7)

La selección de variables se realizó a partir de la revisión de la literatura, consultas a expertos y resultados de los dos estudios pilotos antes mencionados, empleando 25 covariables, relacionadas con la paciente, el tumor, el tratamiento y el diagnóstico, cuya operacionalización se muestra en el Anexo 1.

La recolección del dato primario se realizó mediante la revisión de las historias clínicas de los casos y los controles que integraron la muestra del estudio, los informes de anatomía



patológica y otros documentos con información de las variables en estudio. Para este fin, se confeccionó un formulario que recoge la información necesaria (Anexo 2).

Se aplicaron las 5 primeras etapas previstas en la metodología ASI-IMC: análisis exploratorio, transformación de los datos duplicando la variable dependiente, análisis principal con la variable dependiente como principal y como suplementaria, presentación de los resultados e interpretación de los mismos, según se describe en Sagaró y Zamora.⁸

Para analizar las potencialidades del ASI-IMC en la identificación de posibles causas se decidió comparar la técnica con un criterio externo. En este caso se consideró como la técnica estándar más empleada para estos fines según la literatura, la regresión logística binaria; por lo que se aplicaron ambas técnicas al mismo conjunto de datos y se compararon sus resultados.

Para la aplicación de la regresión logística binaria se verificó el cumplimiento de los supuestos de la prueba, siendo todas las covariables propuestas monótonas, se eliminaron del modelo las variables correlacionadas biológica o estadísticamente y las consideradas intervinientes.

El diagnóstico de la multicolinealidad se realizó a través del factor de inflación de la varianza, creando nuevas combinaciones siempre que fue posible u omitiendo la variable menos importante según la literatura y los expertos. Se comprobó el ajuste del modelo por la prueba de Hosmer y Lemeshow. Se empleó el método por pasos hacia adelante con razón de verosimilitud ya que fue el que mostró un mejor ajuste de los datos al modelo.

Como criterios para el método por pasos, se tomó una probabilidad de entrada de 0,05 y una de eliminación de 0,10; con un punto de corte de 0,5 y un máximo de 20 iteraciones. Se incluyó la constante en el modelo. Se interpretó la significación de los factores a través de la prueba de Wald y del intervalo de confianza de las razones de ventajas representadas por las exponenciales de los coeficientes de regresión ($\text{Exp}[\beta_i]$, donde β_i es el coeficiente asociado a la covariable i -ésima) con un 95% de confianza.

Por último, se compararon ambas técnicas considerando su capacidad de “diagnosticar” factores pronósticos. Dada la respuesta dicotómica de estos procedimientos (identifica o no identifica) se construyó una tabla de contingencia de 2x2, en la cual se consideraron como individuos u objetos de estudio a las covariables empleadas en el mismo y como resultado la identificación de la variable como factor causal o no según ambas técnicas. Se calcularon los indicadores de eficacia de una prueba diagnóstica para el ASI como la sensibilidad, la especificidad, los valores predictivos positivo y negativo y las razones de verosimilitud positiva y negativa, el odds ratio de diagnóstico y los índices Phi, de Youden, de Validez y de Kappa, así como sus intervalos de confianza del 95%, según la expresión de Wilson. El índice de Youden se estimó según su forma clásica propuesta en 1950 y según la versión de Chen de 2015.⁹



Para el cálculo de dichos intervalos de confianza se programó una función en lenguaje R, teniendo en cuenta que se trabajó con un tamaño de muestra pequeño, representado por una cantidad de variables inferior a 30 y que los programas disponibles emplean fórmulas para su cálculo con aproximación a la distribución normal, no aplicable para muestras pequeñas.

Sistemas informáticos y procesadores estadísticos empleados

Para el procesamiento y presentación de los resultados se emplearon los siguientes sistemas:

- IBM SPSS Statistics 24.0, para obtener el modelo de regresión logística binario.
- R 3.2.0 y el RStudio 1.0.44 como entorno de desarrollo integrado sobre el cual se corrió el R para el cálculo del tamaño de muestra y para la programación de las funciones que calculan los indicadores del desempeño del análisis estadístico implicativo comparado con la regresión logística binaria, teniendo en cuenta que todos los programas que calculan los intervalos de confianza de estos indicadores lo hacen con la aproximación a la normal, válido para muestras grandes y en este caso se necesitó la transformación de Wilson.

Resultados

Antes de aplicar la regresión logística se verificaron los supuestos requeridos para su aplicación, arribando a los siguientes resultados:

- No hubo restricciones en la inclusión de covariables en el modelo, pues a pesar de no encontrarse asociación significativa entre algunas de ellas con la variable dependiente, se tomó en consideración su relevancia clínica y el gran tamaño de la muestra. En el anexo 6, (previamente analizado para conocer las frecuencias de las covariables y su asociación con casos y controles) se muestran los resultados de la prueba de homogeneidad y la estimación del riesgo relativo, con un nivel de confianza del 95%.
- Se pudo verificar que todas las covariables propuestas eran monótonas.
- No se conformaron nuevas variables “dummy” por tratarse de variables dicotómicas.

El diagnóstico de multicolinealidad se presenta en la tabla 1, en ella se aprecian los factores de inflación de la varianza (FIV) con valores inferiores a 10 y las tolerancias (T) con valores superiores a 0,1 para las variables individuales y los autovalores (λ) y los índices de condición (η) para cada una de las 25 dimensiones formadas sin incluir la constante en el



modelo (todos inferiores a 30), todo lo cual indica ausencia de multicolinealidad. No se exponen las proporciones de la descomposición de la varianza por no encontrarse colinealidad.

Tabla 1- Diagnóstico de multicolinealidad*

Modelo	Estadísticas de colinealidad		Dimensión	Autovalor	Índice de condición
	Tolerancia	FIV			
Edad	0,358	2,794	1	9,713	1,000
Premenopausia	0,402	2,490	2	1,758	2,351
Tipo histológico	0,219	4,570	3	1,435	2,602
Grado histológico	0,319	3,139	4	1,166	2,886
Grado nuclear	0,263	3,800	5	1,139	2,920
Estadio	0,218	4,580	6	0,989	3,134
Tamaño tumoral	0,475	2,104	7	0,964	3,175
Afectación ganglionar	0,420	2,379	8	0,864	3,352
Metástasis	0,388	2,574	9	0,836	3,408
Invasión sanguínea	0,536	1,865	10	0,755	3,586
Invasión linfática	0,505	1,981	11	0,640	3,896
Invasión perineural	0,620	1,613	12	0,606	4,003
Invasión muscular	0,859	1,164	13	0,500	4,408
Invasión del tejido adiposo	0,508	1,969	14	0,472	4,538
Invasión de piel y pezón	0,858	1,166	15	0,465	4,568
Necrosis	0,388	2,575	16	0,416	4,831
Desmoplasia	0,314	3,187	17	0,401	4,919
Histiocitosis ganglionar	0,331	3,018	18	0,349	5,277
Reacción inflamatoria	0,746	1,340	19	0,308	5,619
Margen quirúrgico libre	0,425	2,351	20	0,291	5,781
Cirugía	0,461	2,169	21	0,224	6,580
Quimioterapia	0,484	2,067	22	0,209	6,815
Radioterapia	0,487	2,053	23	0,191	7,136
Hormonoterapia	0,510	1,960	24	0,164	7,706
Biomarcadores	0,721	1,387	25	0,144	8,209

*Existe multicolinealidad si:

- ✓ $FIV > 10$ $FIV = \frac{1}{1-R_i^2}$, donde R_i^2 es el coeficiente de determinación de la covariable i-ésima con todas las demás.
- ✓ $T > 0,1$; $T = \frac{1}{FVI_i}$
- ✓ $\eta > 30$; $\eta = \sqrt{\frac{\lambda_{max}}{\lambda_i}}$ donde λ_i representa el i-ésimo autovalor y λ_{max} el máximo de los autovalores.

El empleo de la regresión logística, según se aprecia en la tabla 2, determinó como factores de mal pronóstico: un estadio avanzado de la enfermedad, la presencia de metástasis y la quimioterapia y como factores de buen pronóstico los biomarcadores.



Tabla 2- Variables incluidas en el 4to paso de la regresión logística binaria y sus indicadores (Método hacia adelante Wald)

	B	Error estándar	Wald	gl	Sig.	Exp(B)	IC 95%	
							Inferior	Superior
Paso 4 ^d								
Estadio	1,174	,363	10,442	1	,001	3,235	1,587	6,594
Metástasis	2,160	,484	19,898	1	,000	8,673	3,357	22,407
Quimioterapia	1,173	,354	10,963	1	,001	3,230	1,614	6,467
Biomarcadores	-3,172	,502	39,861	1	,000	,042	,016	,112
Constante	-,787	,243	10,458	1	,001	,455		

IC: Intervalo de confianza

Las pacientes que se diagnosticaron en estadio avanzado presentaron una probabilidad de fallecer entre 1,59 y 6,59 veces mayor que las que estaban en los primeros estadios. Las pacientes con metástasis presentaron una probabilidad de fallecer, aproximadamente, entre 3,36 y 22,41 veces mayores que las que no tenían metástasis. De igual forma las que recibieron quimioterapia tuvieron entre 1,61 y 6,47 más probabilidades de morir que las que no recibieron dicho tratamiento.

En cambio, la probabilidad de morir disminuyó entre 0,02 y 0,11 veces en las mujeres tratadas según los resultados de los biomarcadores tumorales con respecto a las que no se les realizó esta determinación.

El valor de la constante en el modelo de regresión logística indica que, de no estar presente ninguno de los factores analizados, la probabilidad de sobrevivir es 2,20 veces (1/ 0,455) mayor que la de fallecer. Las salidas del SPSS para la aplicación de la regresión logística se pueden apreciar en el anexo 4.

Análisis comparativo entre ambas técnicas

Ambas técnicas identificaron como factores pronósticos el estadio, la metástasis, la quimioterapia y los biomarcadores, este último de buen pronóstico, el análisis estadístico implicativo identificó, además, el tamaño tumoral como factor de mal pronóstico. Con estos datos se conformó la tabla de 2x2 que se muestra (Tabla 3).



Tabla 3-Variables identificadas como factores pronóstico de cáncer de mama.

ASI	RL		Total
	Si	No	
Si	4	1	5
No	0	20	20
Total	4	21	25

ASI: Análisis estadístico implicativo

RL: Regresión logística

A partir de los datos anteriores fue posible evaluar el desempeño del análisis estadístico implicativo, cuyos indicadores se muestran en la tabla 4 y su interpretación se comenta a continuación.

Al evaluar la técnica del análisis estadístico implicativo como identificador de factores pronósticos en el cáncer de mama, de acuerdo a los indicadores estimados con un 95 % de confianza, se encontró una capacidad para detectar las variables significativamente asociadas al cáncer entre un 51,01 y un 100 %, ya sea como factor de buen o de mal pronóstico.

También demostró una capacidad entre un 77,33 y un 99,15 % para identificar las variables que no constituyen factores pronósticos. Un factor pronóstico identificado por el análisis estadístico implicativo tiene una probabilidad del 80 % de serlo realmente, así mismo, una variable que no constituye un factor pronóstico según esta técnica, tiene una probabilidad del 100 % de no serlo en realidad. La técnica arrojó un porcentaje de falsos positivos entre 0,85 y 22,67 y de falsos negativos entre 0 y 48,99.

El análisis estadístico implicativo clasificó correctamente entre el 80,46 y 99,29 % de las variables. Con esta técnica es 13,20 veces más probable detectar un factor pronóstico dentro de los verdaderos factores pronósticos que dentro de los que no lo son. Así mismo, es 0.11 veces más probable descartar que un factor sea pronóstico cuando realmente lo es, que descartarlo cuando no lo es. La ganancia promedio de certidumbre neta en la clasificación de un factor pronóstico fue de 0,80. La concordancia entre ambas técnicas y la magnitud de la asociación fueron elevadas (0,86 y 0,87, respectivamente). Con el análisis estadístico implicativo, es 123 veces mayor la posibilidad de identificar un factor como factor pronóstico con respecto a no identificarlo cuando realmente lo es, que la posibilidad de que esto ocurra cuando realmente no lo es.



Tabla 4-Indicadores del desempeño del análisis estadístico implicativo en la identificación de factores pronósticos

Indicador	Valor	Varianza	Intervalo de Confianza	
			Inferior	Superior
Sensibilidad (%)	100,00	0,015619	51,01	100,00
Especificidad (%)	95,24	0,022519	77,33	99,15
Valor Predictivo Positivo (%)	80,00	0,001690	37,55	96,38
Valor Predictivo Negativo (%)	100,00	0,001690	83,89	100,00
Tasa de Falsos positivos (%)	4,76	0,003100	0,85	22,67
Tasa de Falsos negativos (%)	0,00	0,015619	0,00	48,99
Índice de validez o exactitud (%)	96,00	0,002309	80,46	99,29
Razón de verosimilitud positiva	13,20	0,643434	2,74	63,59
Razón de verosimilitud negativa	0,11	1,803326	0,01	1,49
Índice de Youden	0,80	0,062500	0,46	1,44
Índice de Youden (según Chen)	0,80	0,032000	0,60	1,30
Índice de Kappa	0,86	0,438276	-0,43	2,16
Coficiente Phi	0,87	1,250000	-1,32	3,06
Logaritmo del odds ratio de diagnóstico	4,81	2,937669	1,45	8,17
Odds ratio de diagnóstico	123,00	---	4,27	3538,61

Discusión

La presencia de metástasis y el avanzado estadio de la enfermedad también fueron identificados como factores de mal pronóstico en el estudio de Pardo,⁷ no así en el de Moraga,⁶ lo cual pudo estar condicionado por un tamaño de muestra muy inferior al de estos estudios.

En otros estudios como el de Maffuz de 2017, sobre mujeres mexicanas, la mayoría de las diagnosticadas con cáncer de mama presentaron estadios avanzados al momento del diagnóstico (52,9 % de las pacientes).¹⁰

La metástasis es uno de los factores de mal pronóstico más evidenciado en muchas investigaciones, entre las que figuran la de Kim que encontró como el principal factor pronóstico la metástasis en ganglios linfáticos, la de Chávez que estudia la supervivencia de las pacientes con cáncer de mama metastásico y las de Mohar-Betancourt y Wiltzel que valoran la metástasis cerebral.⁽¹¹⁻¹⁴⁾



Los estudios de Kim y Grassadonia encontraron la quimioterapia neoadyuvante como factor de buen y mal pronóstico, respectivamente, coincidiendo el último con este estudio. ^(15, 16)

La regresión, a diferencia de la metodología ASI-IMC no identificó el tamaño del tumor como factor pronóstico, aunque este es bien reconocido como uno de los más importantes factores pronósticos, sobre todo, en el momento del diagnóstico, es por ello que las pesquisas con mamografía han permitido detectar tumores incipientes, prolongándose la supervivencia de las mujeres que padecen la enfermedad. ^(17, 18)

Sobre el análisis comparativo entre la regresión logística y el análisis estadístico implicativo, en el presente estudio ambas técnicas logran reconocer los mismos factores de mal pronóstico, pues la quimioterapia también es reconocida por el análisis estadístico implicativo cuando se baja el índice de implicación hasta un 89%, también coinciden en uno de los factores de buen pronóstico. Los indicadores estimados mostraron un desempeño adecuado de la metodología, lo cual coincide con Pardo y Moraga, ^(5,6) quienes analizaron 22 covariables, casi todas similares, pero mostraron limitaciones con respecto a la identificación de los biomarcadores, lo que es superado en este estudio donde se trabajó con 25 covariables. Pardo detectó una mayor sensibilidad y menor especificidad del análisis estadístico implicativo que Moraga y una mayor eficacia en general. ^(5,6) El resto de los estudios que emplearon la metodología propuesta con el mismo objetivo, pero en otro tipo de enfermedades también arrojaron resultados aceptables de la eficacia de la misma, tanto para la identificación de factores de riesgo, ¹⁹ como para la identificación de factores pronósticos de mortalidad por cáncer, ⁽²⁰⁻²⁴⁾ o pronósticos de otras enfermedades. ^(25,26)

Se concluye el estudio con una evaluación satisfactoria de la efectividad de la metodología de aplicación del ASI en la identificación de posibles factores causales en las investigaciones en salud y a la vez fueron identificados los factores pronósticos de mortalidad por cáncer de mama en la población estudiada.

Referencias

1. Sagaró Del Campo NM, Zamora Matamoros L. Evolución histórica de las técnicas estadísticas y las metodologías para el estudio de la causalidad en ciencias médicas. MEDISAN. 2019 [citado 11Mar2021];23(3):[aprox. -556 p.]. Disponible en: <http://www.medisan.sld.cu/index.php/san/article/view/2434>
2. Gras R, Régnier JC, Lahanier-Reuter D, Marinica C, Guillet F. L'Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales 3ème édition revue et augmentée. Toulouse: Cépaduès Ed; 2017.



3. Sagaró Del Campo NM, Zamora Matamoros L. ¿Por qué emplear el análisis estadístico implicativo en los estudios de causalidad en salud? Revista Cubana de Informática Médica 2019[citado 11 Mar 2021];11(1):88-103. Disponible en: http://revinformatica.sld.cu/index.php/rcim/article/view/316/pdf_94
4. Sagaró Del Campo NM, Zamora Matamoros L. Propuesta metodológica de contextualización del análisis estadístico implicativo a las investigaciones médicas de causalidad. RCIM. 2020[citado 11 Mar 2021];12(1):31-43. Disponible en:
5. Rodríguez Sarría Y.M, Delisle Ureña G, Sagaró del Campo N.M, Escalona Fonseca M. Factores pronósticos y supervivencia de mujeres con cáncer de mama en Santiago de Cuba MEDISAN 2018[citado 11 Mar 2021];22(5):485. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192018000500004
6. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de mama. MEDISAN. 2017[citado 11 Mar 2021];21(4): 413. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192017000400003
7. Pardo-Santana S, Sagaró-del-Campo NM, Zamora-Matamoros L, Viltre-Castellanos DM. Utilidad del análisis estadístico implicativo para identificar factores pronósticos en pacientes con cáncer de mama. Revista Electrónica Dr. Zoilo E. MarinelloVidaurreta [revista en Internet]. 2019 [citado 11 Mar 2021];44(4). Disponible en: <http://revzoilomarinaldo.sld.cu/index.php/zmv/article/view/1869>
8. Sagaró Del Campo NM, Zamora Matamoros L, BonneBartutis MT. Identificación de posibles factores causales en salud aplicando la metodología “ASI-IMC”. Multimed. 2021[citado 11 Jun 2021];25(4). Disponible en: <http://www.revmultimed.sld.cu/index.php/mtm/article/view/1910>
9. Sagaró Del Campo NM, Zamora Matamoros L. ¿Cómo se creó una metodología para aplicar el análisis estadístico implicativo en las investigaciones clínicas? Gac. Méd. Espirit. 2020[citado 11 Mar 2021];22(1):15-36. Disponible en:<http://scielo.sld.cu/pdf/gme/v22n1/1608-8921-gme-22-01-15.pdf>
10. MaffuzAziz A, Labastida Almendaro S, Espejo Fonseca A, Rodríguez Cuevas S. Características clinicopatológicas del cáncer de mama en una población de mujeres en México. Cirugía y Cirujanos. 2017[citado 11 Mar 2021];85(3):201-7. Disponible en: <https://www.redalyc.org/pdf/662/66250860003.pdf>.
11. Kim JM, Kim JY, Jung EJ, Kwag SJ, Park JH, Park T, et al. The prognosis factors among breast cancer patients with extensive axillary lymph node metastasis. Korean J ClinOncol. 2018[citado 11 Mar 2021];14(1):43-7. Disponible en: <https://www.koreamed.org/article/0247KJCO/2018.14.1.43>
12. Chávez Mac Gregor M, Hortobagyi GN. Avances en el manejo del Cáncer de Mama Metastásico. ¿Existen subgrupos curables? En: Cáncer De Mama: Aspectos de Interés Actual. Fundación de Estudios Mastológicos; 2012. Cap 20.



13. 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 PublicaMex. 2018[citado 11 Mar 2021];60(2):141-50. Disponible en: <http://saludpublica.mx/index.php/spm/article/view/9082/11453>
14. Wiltzel I, Olivera-Ferrer L, Pantel K, Müller V, Wikman H. Breast cancer brain metastases: biology and new clinical perspectives. BreastCancer Res. 2016[citado 11 Mar 2021];18(1):8. Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4717619/pdf/13058_2015_Article_665.pdf
15. Kim HS, Yoo TK, Park WC, Chae BJ. Potential Benefits of Neoadjuvant Chemotherapy in Clinically Node-Positive Luminal Subtype– Breast Cancer. J BreastCancer. 2019[citado 11 Mar 2021];22(3):412-424 Disponible en: <https://pubmed.ncbi.nlm.nih.gov/31598341/>
16. Grassadonia A, Vici P, Gamucci T, Moscetti L, Pizzuti L, Mentuccia L, et al. Long-term outcome of breast cancer patients with pathologic N3a lymph node stage. Breast 2017[citado 11 Mar 2021];32:79-86. Disponible en: https://www.clinicalkey.es/service/content/pdf/watermarked/1-s2.0-S0960977616302879.pdf?locale=es_ES&searchIndex=
17. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. N Engl J Med 2016[citado 11 Mar 2021];375:1438-47. Disponible en: <https://www.ncbi.nlm.nih.gov/pubmed/27732805>
18. Adrover Cebrián E, SabadelMercadal MD, Izquierdo Sanz M, Prag de Puig M, Modolell Roig A. Factores pronósticos y predictivos en cáncer de mama. Una visión evolutiva de la morfología a la genética. España:Sociedad Española de Patología Mamaria; 2017. [citado 11 Mar 2021]. Disponible en: <https://docplayer.es/95738177-Factores-pronosticos-y-predictivos-en-cancer-de-mama.html>
19. García Mederos Y, Zamora Matamoros L, Sagaró del Campo N. Análisis estadístico implicativo en la identificación de factores de riesgo en pacientes con cáncer de pulmón. MEDISAN. 2015;19(8). [citado 11 Mar 2021]; Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192015000800003
20. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de pulmón. MEDISAN. 2016[citado 11 Mar 2021];20(3). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192016000300010
21. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de próstata. MEDISAN. 2018[citado 11 Mar 2021];22(1). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192018000100007



22. Galano Vázquez K, Sagaró del Campo NM, Zamora Matamoros L, Lambert Matos Y, Mingui Carbonell E. Análisis estadístico implicativo en la identificación de factores pronósticos de mortalidad del cáncer renal. Rev. inf. cient. [Internet]. 2018 [citado 11 Mar 2021];97(4). Disponible en: <http://www.revinfcientifica.sld.cu/index.php/ric/article/view/2268>
23. Sagaró-del-Campo N, Zamora-Matamoros L. Análisis estadístico implicativo en la identificación de factores pronósticos de mortalidad por cáncer cervicouterino. *Acta Médica del Centro*. 2021[citado 11 Mar 2021];15(2). Disponible en: <http://www.revactamedicacentro.sld.cu/index.php/amc/issue/view/53>
24. Lamber-Matos Y, Sagaró-del-Campo N, Zamora-Matamoros L. Identificación de factores pronósticos en cáncer cérvico-uterino mediante análisis estadístico implicativo. Archméd Camagüey. 2021[citado 11 Mar 2021];25(4):e7956. Disponible en: <http://revistaamc.sld.cu/index.php/amc/view/7956>
25. Rodríguez-Pérez I, Sagaró-del-Campo NM, Zamora-Matamoros L, Martínez-Álvarez A. Factores pronósticos en la enfermedad pulmonar obstructiva crónica exacerbada. Rev. electron. Zoilo. 2021[citado 11 Mar 2021];46(2). Disponible en: <http://revzoilomarinaldo.sld.cu/index.php/zmv/article/view/2643>
26. Paez Candelaria Y, Sagaró del Campo NM, Zamora Matamoros L. Análisis estadístico implicativo en la determinación de factores pronósticos del estado nutricional del paciente grave al egreso. MEDISAN 2018[citado 11 Mar 2021];22(6). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192018000600007



ANEXO 1-Operacionalización de las variables para la identificación de factores pronósticos de mortalidad por cáncer de mama.

Variable (Nombre en la base de datos)	Categoría	Definición operacional de las categorías
Variable de respuesta:		
Estado	Viva Fallecida	Según el estado vivo o fallecido por causa de la enfermedad al término del estudio.
Variables explicativas:		
<ul style="list-style-type: none"> Factores dependientes del paciente 		
1. Mayor de 50 años (EDAD)	Si No	Según tenga o no más de 50 años cumplidos al momento del diagnóstico.
2. Premenopausia (Menopau)	Si No	Según se realizó o no el diagnóstico en la etapa premenopáusica.
<ul style="list-style-type: none"> Factores dependientes del tumor 		
3. Tipo histológico (TipoHistol)	Peor Mejor	Los tipos ductal infiltrante, lobulillar, mixto corresponden a la categoría peor, y los tipos papilar, medular o mucinoso a la mejor.
4. Grado de diferenciación histológica (GradoHistol)	Peor Mejor	Según la modificación de Elston-Ellis del sistema de graduación Bloom-Richardson ^{A,B} se consideraron los de mayor grado (pobremente diferenciados) como peor y los grados I y II (bien y moderadamente diferenciados) como mejor.
5. Grado de diferenciación nuclear (GradoNucle)	Peor Mejor	Los grados III, IV corresponden a peor y I, II al mejor.
6. Estadio (Estadio)	Peor Mejor	De acuerdo a la clasificación TNM de la 7ma. Edición de la American Joint Committee on Cancer (AJCC) ^C se consideraron como: <ul style="list-style-type: none"> peor los estadios siguientes: <ul style="list-style-type: none"> IIIA (T₀N₂M₀ o T₁N₂M₀ o T₂N₂M₀ o T₃N₁M₀ o T₃N₂M₀), IIIB (T₄N₀M₀ o T₄N₁M₀ o T₄N₂M₀), IIIC (Cualquier T y N₃M₀), IV (Cualquier T, cualquier N y M₁). y mejor: <ul style="list-style-type: none"> 0 (T₁N₀M₀), I (T₁N₀M₀), IIA (T₀N₁M₀ o T₁N₁M₀ o T₂N₀M₀), IIB (T₂N₁M₀ o T₃N₀M₀).
7. Tamaño tumoral (Tamaño)	Peor Mejor	Se consideró peor un tumor mayor o igual a 5 cm (T ₃ o T ₄) y mejor uno menor de 5 cm (T ₀ , T ₂ , T ₁ o T ₂).
8. Afectación ganglionar (Ganglio)	Si No	Según la presencia (N ₁ a N ₃) o no (N ₀) de metástasis de los ganglios linfáticos regionales.
9. Metástasis (Meta)	Si No	Según la evidencia (M ₁) o no (M ₀) de metástasis a distancia.
10. Invasión sanguínea (InvSangre)	Si No	Según presente o no invasión sanguínea.
11. Invasión linfática (InvLinfa)	Si No	Según presente o no invasión linfática.



12. Invasión perineural (InvNeura)	Si No	Según presente o no invasión perineural.
13. Invasión muscular (InvMuscul)	Si No	Según presente o no invasión muscular.
14. Invasión del tejido adiposo (InvAdipo)	Si No	Según presente o no invasión del tejido adiposo.
15. Invasión de piel y pezón (PielPezón)	Si No	Según presente o no invasión de piel y pezón.
16. Necrosis (Necrosis)	Si No	Según presente o no necrosis.
17. Desmoplasia (Desmopla)	Si No	Según presente o no desmoplasia.
18. Histiocitosis ganglionar (Histiocito)	Si No	Según presente o no histiocitosis ganglionar.
19. Reacción inflamatoria (ReacInfla)	Si No	Según presente o no una reacción inflamatoria.
20. Margen quirúrgico libre (Margen)	Si No	Según presente o no distancia entre el borde del tumor y la tinta china.
<ul style="list-style-type: none"> Factores dependientes del tratamiento 		
21. Cirugía (Cirugia)	Si No	Según reciba o no tratamiento quirúrgico.
22. Quimioterapia (Quimio)	Si No	Según reciba o no quimioterapia.
23. Radioterapia (Radio)	Si No	Según reciba o no radioterapia.
24. Hormonoterapia (Hormona)	Si No	Según reciba o no hormonoterapia.
<ul style="list-style-type: none"> Factores dependientes del diagnóstico 		
25. Biomarcadores (Biomar)	Si No	Según se hayan obtenido o no al diagnóstico el resultado de los biomarcadores tumorales.

- Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer. 1957[citado 5 sep 2019];11:359-77. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2073885/pdf/brjancer00376-0041.pdf>
- Elston CW. Grading of invasive carcinoma of the breast En: Page DL, Anderson TJ eds. Diagnostic Histopathology of the Breast. 1st ed. Edinburgh: Churchill Livingstone; 1987. pp 300-11.
- Edge S, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, AJCC. Cancer Staging Manual, 7th Ed, New York; 2009.

Codificación en la base de datos:

Se codificó como 1 a las fallecidas, los “Si “y los “Peor”, 0 en caso contrario,



ANEXO 2-Formulario de recogida de datos

Estudio de factores pronósticos de mortalidad en cáncer de mama

HC _____

Viva () Fallecida ()

1. Mayor de 50 años Si () No ()
2. Premenopausia Si () No ()
3. Tipo histológico _____
4. Grado de diferenciación histológica I () II () III () IV ()
5. Grado de diferenciación nuclear I () II () III () IV ()
6. Estadio (clasificación TNM) 0 () I () II () III () IV ()
7. Tamaño tumor primario < 5 () ≥ 5 ()
8. Afectación ganglionar Si () No ()
9. Metástasis Si () No ()
10. Invasión sanguínea Si () No ()
11. Invasión linfática Si () No ()
12. Invasión perineural Si () No ()
13. Invasión muscular Si () No ()
14. Invasión adiposa Si () No ()
15. Invasión de piel y pezón Si () No ()
16. Necrosis Si () No ()
17. Desmoplasia Si () No ()
18. Histiocitosis ganglionar Si () No ()
19. Reacción inflamatoria Si () No ()
20. Margen quirúrgico libre Si () No ()
21. Cirugía Si () No ()
22. Quimioterapia Si () No ()
23. Radioterapia Si () No ()
24. Hormonoterapia Si () No ()
25. Biomarcadores Si () No ()



ANEXO 3-Descripción de casos y controles**Tabla**-Casos y controles según la presencia de factores de mal pronóstico y significación de la asociación.

Factor Pronóstico	Casos		Controles		p*
	No.	%	No.	%	
Mayor de 50 años	29	20,71	33	23,57	0,666
Premenopausia	20	14,29	23	16,43	0,740
Tipo histológico	120	85,71	107	76,43	0,067
Grado de diferenciación histológica	36	25,71	33	23,57	0,782
Grado de diferenciación nuclear	59	42,14	42	30,00	0,046
Estadio	97	69,29	38	27,14	0,000
Tamaño tumoral	41	29,29	16	11,43	0,000
Afectación ganglionar	71	50,71	48	34,29	0,008
Metástasis	73	52,14	15	10,71	0,000
Invasión sanguínea	61	43,57	50	35,71	0,222
Invasión linfática	61	43,57	59	42,14	0,904
Invasión perineural	37	26,43	21	15,00	0,027
Invasión muscular	5	3,57	3	2,14	0,720
Invasión del tejido adiposo	40	28,57	52	37,14	0,162
Invasión de piel y pezón	5	3,57	6	4,29	1,000
Necrosis	73	52,14	70	50,00	0,811
Desmoplasia	79	56,43	78	55,71	1,000
Histiocitosis ganglionar	67	47,86	64	45,71	0,811
Reacción inflamatoria	18	12,86	16	11,43	0,855
Margen quirúrgico libre	78	55,71	61	43,57	0,056
Cirugía	37	26,43	30	21,43	0,401
Quimioterapia	57	40,71	29	20,71	0,000
Radioterapia	46	32,86	49	35,00	0,801
Hormonoterapia	63	45,00	67	47,86	0,719
Biomarcadores	9	6,43	67	47,86	0,000

*Chi cuadrado de Pearson con corrección de Yates



ANEXO 4- Salida del SPSS para la regresión logística binaria

VARIABLES INCLUIDAS EN EL 4TO PASO DE LA REGRESIÓN LOGÍSTICA BINARIA Y SUS INDICADORES
(Método hacia adelante Wald)

	B	Error estándar	Wald	gl	Sig.	Exp(B)	IC 95%	
							Inferior	Superior
Paso 4^d								
Estadio	1,174	,363	10,442	1	,001	3,235	1,587	6,594
Metástasis	2,160	,484	19,898	1	,000	8,673	3,357	22,407
Quimioterapia	1,173	,354	10,963	1	,001	3,230	1,614	6,467
Biomarcadores	-3,172	,502	39,861	1	,000	,042	,016	,112
Constante	-,787	,243	10,458	1	,001	,455		

Pruebas ómnibus de coeficientes de modelo

		Chi-cuadrado	gl	Sig,
Paso 4	Paso	10,814	1	,001
	Bloque	149,095	4	,000
	Modelo	149,095	4	,000

Resumen del modelo

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
4	239,068 ^b	,413	,550

b, La estimación ha terminado en el número de iteración 6 porque las estimaciones de parámetro han cambiado en menos de ,001.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig,
4	2,945	6	,816



Tabla de contingencia para la prueba de Hosmer y Lemeshow

		ESTADO = Viva		ESTADO = Fallecida		Total
		Observado	Esperado	Observado	Esperado	
Paso 4	1	35	35,326	1	,674	36
	2	23	23,696	3	2,304	26
	3	41	42,601	21	19,399	62
	4	22	18,653	18	21,347	40
	5	13	11,726	16	17,274	29
	6	4	4,430	17	16,570	21
	7	2	2,976	39	38,024	41
	8	0	,591	25	24,409	25

Tabla de clasificación^a

		Pronosticado			Porcentaje correcto
		ESTADO			
Paso 4	Observado	Viva	Fallecida		
	ESTADO	Viva	108	32	77,1
Fallecida		26	114	81,4	
Porcentaje global				79,3	

a, El valor de corte es ,500

		Variables en la ecuación					95% C.I. para EXP(B)		
		B	Error estándar	Wald	gl	Sig,	Exp(B)	Inferior	Superior
Paso 4 ^d	Estadio	1,174	,363	10,442	1	,001	3,235	1,587	6,594
	Metástasis	2,160	,484	19,898	1	,000	8,673	3,357	22,407
	Quimioterapia	1,173	,354	10,963	1	,001	3,230	1,614	6,467
	Biomarcadores	-3,172	,502	39,861	1	,000	,042	,016	,112
	Constante	-,787	,243	10,458	1	,001	,455		

d, Variables especificadas en el paso 4: Estadio,



Conflictos de interés

Las autoras declaran no tener conflicto de intereses y que no han recibido financiamiento para la investigación.

Contribución de autoría

Dra. C. Nelsa María Sagaró del Campo y Dra. C. Larisa Zamora Matamoros: Ambas desarrollaron la conceptualización, curación de datos, análisis formal, investigación, metodología, administración del proyecto, supervisión, validación, visualización, redacción, revisión y edición del borrador y del original.

