

Tendencias y comportamientos de los datos disponibles sobre el cáncer de mama

Trends and behaviors in available data on breast cancer

Guillermo Molero Castillo

Doctor en Tecnologías de Información, Consejo Nacional de Ciencia y Tecnología (CONACYT-Universidad Veracruzana), México D.F., México

Tania Pilar Álvarez Domínguez, Médica Cirujana, Universidad Nacional Autónoma de México; México D.F., México

Yaimara Céspedes González, Doctora en Ciencias Administrativas y Gestión para el Desarrollo, Universidad Veracruzana, México D.F., México

Acceda a este artículo en siicsalud

www.siicsalud.com/dato/experto.php/155526

Recepción: 14/9/2017 - Aprobación: 2/7/2018
Primera edición: 25/7/2018

Enviar correspondencia a: Guillermo Molero Castillo, Av. Insurgentes Sur 1582, Col. Crédito Constructor, Del. Benito Juárez, 03940, México D.F., México
ggmolero@conacyt.mx



Especialidades médicas relacionadas, producción bibliográfica y referencias profesionales de los autores



www.dx.doi.org/10.21840/siic/155526

Abstract

Data analysis aims to discover information of interest in diverse application fields, such as health, in which studies can be made for the prognosis of trends and behaviors that occur in certain pathologies, such as cancer. Objective: This study aimed to present the analysis and characterization of breast cancer cases diagnosed in Hispanic-origin women. Method: A qualitative and quantitative approach was used to conduct this study. The data source used corresponds to clinical records of the Surveillance, Epidemiology, and End Results program of the National Cancer Institute of the United States. The analysis was performed in two stages. The first consisted of a preliminary analysis of the total variables listed in the database, with the purpose of establishing those relevant variables, according to the period of their registers. In the second stage, the quality of the data series was determined to establish the significant variables associated with the diagnosis of breast cancer in Hispanic-origin women. Results: We analyzed the variability and distribution of the main, previously identified, oncological variables that record specific information about the patient and breast cancer. Conclusion: We observed that breast cancer is a disease with non-uniform behavior, indicating that the possibility of recovery depends on the stage of cancer, whether it is located exactly in the breast or has spread to other parts of the body, type, treatment received, and the health of the patient in general.

Key words: data analysis, time series studies, breast neoplasms, women's health

Resumen

El análisis de datos es útil para descubrir información de interés en diversos campos de aplicación, como en salud, donde se pueden realizar estudios para el pronóstico de tendencias y comportamientos que presentan determinadas enfermedades, como el cáncer. *Objetivo:* En este trabajo se presenta el análisis y la caracterización de casos de cáncer de mama en mujeres de origen hispano. *Método:* Para realizar este estudio se empleó un enfoque cualitativo y cuantitativo. La fuente de datos utilizada corresponde a registros clínicos de la base de datos del programa de Vigilancia, Epidemiología y Resultados Finales del Instituto Nacional del Cáncer de los Estados Unidos. El análisis se realizó en dos etapas. La primera consistió en un análisis preliminar del total de variables incluidas en la base de datos con el propósito de establecer las variables relevantes, según el periodo de sus registros. En la segunda etapa se determinó la calidad de la serie de datos para establecer las variables significativas asociadas con el diagnóstico de cáncer de mama en mujeres de origen hispano. *Resultados:* Se analizó la variabilidad y distribución de las principales variables oncológicas, previamente identificadas, que registran información puntual sobre el paciente y el cáncer de mama. *Conclusión:* Se observó que el cáncer de mama es una enfermedad con un comportamiento no uniforme, lo que indica que la posibilidad de recuperación depende de la etapa del cáncer, de si se encuentra exclusivamente en la mama o se ha diseminado a otras partes del cuerpo, así como del tipo, el tratamiento recibido y la salud de la paciente en general.

Palabras clave: análisis de datos, estudios de series temporales, neoplasias de la mama, salud de la mujer

Introducción

En la última década se ha observado un incremento en la aplicación del análisis de datos a problemas relacionados con series de tiempo.¹ Dichos trabajos han sido orientados principalmente al agrupamiento, para el análisis de la causa de muerte; la clasificación, enfocado a la predicción del consumo de fármacos; la detección de anomalías, orientado al análisis de historias clínicas para la identificación de enfermedades, y el descubrimiento de reglas, dirigido a la identificación de distintas enfermedades.

Las series de datos temporales son un caso particular de patrones secuenciales, su análisis ofrece una valoración de la estacionalidad de la serie, describiendo las oscilacio-

nes de los datos con relación a un valor promedio e identificando la presencia de posibles tendencias.² Se define como serie de datos temporales al conjunto de valores ordenados cronológicamente que permiten describir y predecir el comportamiento de una o más variables en un determinado periodo.³ Algunas veces estas series pueden ser extensas, y contener billones de observaciones.⁴

En las series temporales se identifican cuatro tipos de patrones: tendencia, variación estacional, variación accidental y variación cíclica.⁵ La tendencia (T) refleja la evolución de la serie durante un determinado periodo. Este periodo varía según la naturaleza de la serie, el cual puede ser estacionario o constante, lineal, exponencial o de otro tipo. La variación estacional (S) es el compor-

tamiento que agrupa las oscilaciones repetitivas en períodos. Estos períodos pueden ser estaciones del año, días, meses, bimestres, trimestres, semestres, años, entre otros. Mientras que la variación accidental (A) es un patrón que corresponde a las fluctuaciones accidentales que se dan por la aparición de fenómenos imprevisibles, como la presencia de huracanes, que afectan la variable en estudio de manera esporádica y no permanente. También es conocida como variación irregular. Por otra parte, la variación cíclica (C) se presenta cuando los datos reflejan oscilaciones periódicas no regulares, ocasionadas por asumir períodos no establecidos. Generalmente aparecen en series de datos climatológicos, por ejemplo, en ciclos de sequía.

En este trabajo se describe el análisis y la caracterización de la fuente de datos relacionada con el cáncer de mama en mujeres de origen hispano, residentes en los Estados Unidos. Los datos analizados corresponden a series de datos del programa de Vigilancia, Epidemiología y Resultados Finales (SEER) del Instituto Nacional del Cáncer (NCI, por sus siglas en inglés) de los Estados Unidos. La principal consideración fue determinar cuántas y cuáles son las variables oncológicas apropiadas para el estudio. Se analizó también la variabilidad y distribución de las principales variables identificadas en la base de datos SEER, como origen del paciente, edad, año de diagnóstico, tipo de enfermedad y confirmación del diagnóstico.

Método

Fuente de datos

La fuente de datos utilizada corresponde a registros clínicos de la base de datos del programa SEER. Este programa empezó a recopilar y registrar datos sobre los diferentes tipos de cáncer a partir de 1973, actualmente es responsable en los Estados Unidos del registro nacional de cáncer y la principal fuente de información autorizada para esta enfermedad. El SEER recolecta información sobre los casos de cáncer diagnosticados (incidencia), las muertes atribuidas a esta enfermedad (mortalidad) y la supervivencia de pacientes con cáncer. Esto con el fin de comprender y abordar el cáncer en la población de los Estados Unidos. En la actualidad, son diversas las investigaciones que se realizan a través de estos registros del cáncer, los cuales están a disposición de investigadores, médicos, funcionarios de salud pública, políticos, grupos de investigación y público en general.⁶ Esto con el fin de monitorizar las tendencias del cáncer con el paso del tiempo, apoyar en el establecimiento de prioridades en la asignación de recursos, guiar la planificación y evaluación de programas para el control del cáncer y promover actividades de investigación en el área médica y de epidemiología.

Así, la información sobre casos de cáncer y muertes por esta enfermedad es útil para elaborar informes sobre las tendencias del cáncer, determinar si los esfuerzos de prevención y control son eficaces, propiciar la participación en investigaciones y emprender acciones cuando se informen posibles aumentos en la incidencia del cáncer.

Base de datos SEER

La base de datos SEER comenzó a reunir y registrar datos a partir de 1973 sobre los diversos tipos de cáncer en Connecticut, Iowa, Nuevo Mexico, Utah, Hawai y áreas metropolitanas de Detroit, San Francisco y Oakland.⁷ En

los últimos treinta años, SEER ha añadido más poblaciones a la lista de vigilancia y ahora existen millones de casos registrados en la base de datos. Abarca, en la actualidad, aproximadamente el 28% de la población de los Estados Unidos.⁸

La base de datos contiene datos demográficos del paciente, localización del tumor primario, morfología del tumor, etapa del cáncer al momento del diagnóstico, tratamiento, seguimiento de la enfermedad, entre otros. La obtención y el registro de los datos se realiza a través de establecimientos médicos, como hospitales, consultorios y laboratorios de patología, que envían información sobre los casos evaluados a sus respectivos registros estatales de cáncer.⁹ Por lo general, la mayor parte de la información proviene de hospitales, donde empleados autorizados transfieren la información de las historias clínicas de los pacientes a bases de datos locales, para posteriormente ser enviadas al registro central del cáncer.¹⁰

Análisis de la fuente de datos

El análisis de la fuente de datos se realizó en dos etapas. La primera consistió en una revisión y análisis preliminar del total de variables incluidas en la base de datos SEER, con el fin de establecer aquellas variables relevantes en función del período de sus registros, descartando las que presentaron escasa o nula cantidad de registros disponibles. La base de datos contaba con registros de 1973 a 2008, por lo que se hizo el análisis a partir de esa fecha. En la segunda etapa se determinó la calidad de la serie de datos para establecer las variables significativas del diagnóstico de mujeres de origen hispano con cáncer de mama.

Análisis preliminar

Como parte del análisis preliminar se identificaron las variables con suficientes registros, esto es que tuvieran al menos más del 50% de datos a lo largo de 1973-2008. Se hizo esto con la finalidad de tener variables con períodos similares, descartando las que presentan una alta cantidad de registros faltantes. Para analizar la cantidad de datos en las variables disponibles (124 campos), se organizaron los archivos de la fuente de datos de texto plano en un archivo único. Esto con el propósito de importar los datos (filas y columnas) en una tabla de la base de datos en SQL Server.

Para la separación de las variables en la tabla se utilizó la información proporcionada por el programa SEER sobre el nombre de la variable, la posición y la longitud. Para esto, se elaboró un procedimiento en SQL, formando así la tabla con 124 campos y 1 041 736 registros. De estos, todos asociados con el cáncer de mama, se filtraron solo los datos de pacientes de origen hispano, por lo que quedaron en total 67 156 registros.

En este sentido, como resultado del análisis preliminar de las 124 variables, se observó que 73 tenían suficientes registros disponibles, es decir, contaban con porcentajes por encima del 50% de datos registrados a lo largo de 1973-2008. Mientras que las otras variables, en total 51, tenían una alta cantidad de registros nulos o faltantes.

Análisis de la calidad de datos

La segunda etapa de análisis consistió en determinar la calidad de la serie de datos, asociada fundamentalmente con el número de registros válidos continuos en un determinado período. Se hizo esto con la finalidad de determi-

nar las variables significativas vinculadas con el diagnóstico de los casos de cáncer de mama en mujeres de origen hispano.

Para este proceso, y sobre la base de los resultados obtenidos en el análisis preliminar, se examinaron los registros disponibles de cada una de las variables, en total, 73. El análisis consistió en seleccionar aquellas variables significativas que tienen relación directa con el cáncer de mama, con registros suficientes en períodos consecutivos y bajo la opinión de oncólogos especialistas del Hospital General Vladimir Ilich Lenin y del Policlínico Pedro Díaz Coello, de Holguín, Cuba. Se buscó la participación de estos especialistas para fortalecer la investigación sobre la incidencia del cáncer de mama en pacientes de origen hispano.⁵ La participación de estos especialistas se logró mediante una colaboración académica y la realización de una estancia de investigación. Así, para el análisis de los datos y la selección de las variables, se establecieron las siguientes consideraciones: la variable debe tener relación directa con el cáncer de mama y no con otros tipos de cáncer registrados por el programa SEER; cada variable debe tener por lo menos cuatro años de datos consecutivos, a partir de 1973; la variable analizada debe tener al menos el 90% de registros válidos consecutivos; la selección de las variables estuvo sujeta a la opinión de médicos especialistas.

Al aplicar las consideraciones mencionadas se observó que 35 variables cumplieron con los criterios establecidos (Tabla 1).

El número de variables seleccionadas representa el 47.9% del total de variables que fueron elegidas mediante el análisis preliminar (73 variables), y el 28.2% del total general establecido en la base de datos SEER (124 variables). Por otra parte, las variables seleccionadas presentan un alto porcentaje de aceptación (100%); por su parte, las variables descartadas no fueron seleccionadas debido a que no tenían relación directa con el cáncer de mama, presentaban redundancia o duplicidad de información con otras variables, y por la opinión de médicos especialistas que colaboraron en esta etapa del trabajo de investigación.

La variable SEX, que registra el sexo del paciente al momento del diagnóstico, que tiene relación directa con el cáncer de mama, también fue descartada. Esto debido a uno de los alcances de este estudio, que es trabajar con registros de casos de cáncer de mama en mujeres de origen hispano. Así, de los 67 156 registros, 312 corresponden a pacientes hispanos de sexo masculino, los cuales fueron descartados, por lo que quedaron en total 66 844 registros.

Tabla 1. Variables seleccionadas.

N°	Variable	Descripción	Año de inicio	Año final	Años	Registros válidos	% válidos
1	CASENUM	Identificador único	1973	2008	35	67 156	100
2	REG	Identificador de población	1973	2008	35	67 156	100
3	MAR_STAT	Estado civil	1973	2008	35	67 156	100
4	ORIGIN	Origen del paciente	1973	2008	35	67 156	100
5	NHIA	Origen del paciente NHIA	1973	2008	35	67 156	100
6	AGE_DX	Edad del paciente	1973	2008	35	67 156	100
7	YR_BIRTH	Año de nacimiento	1973	2008	35	67 152	99.9
8	PLC_BIRTH	Lugar de nacimiento	1973	2008	35	67 156	100
9	SEQ_NUM	Secuencia de tumores	1973	2008	35	67 156	100
10	DATE_mo	Mes de diagnóstico	1973	2008	35	67 156	100
11	DATE_yr	Año de diagnóstico	1973	2008	35	67 156	100
12	LATERAL	Lado del tumor	1973	2008	35	67 156	100
13	BEHO3V	Tipo de neoplasia	1973	2008	35	67 156	100
14	GRADE	Tipo de células cancerígenas	1973	2008	35	67 156	100
15	DX_CONF	Confirmación del cáncer	1973	2008	35	67 156	100
16	REPT_SRC	Fuente de información	1973	2008	35	67 156	100
17	EOD10_SZ	Diámetro del tumor	1988	2003	15	35 698	100
18	EOD10_PN	Ganglios linfáticos examinados	1988	2008	20	63 400	100
19	EOD10_NE	Ganglios linfáticos retirados	1988	2008	20	63 400	100
20	CS_SIZE	Tamaño del tumor	2004	2008	4	27 702	100
21	D_AJCC_S	Etapa de la enfermedad	2004	2008	4	27 702	100
22	SURGRIM	Procedimiento quirúrgico	1998	2008	10	50 672	100
23	NO_SURG	Motivo de la no cirugía	1973	2008	35	67 156	100
24	RADIATN	Método de radioterapia	1973	2008	35	67 156	100
25	RAD_SURG	Secuencia cirugía y radioterapia	1973	2008	35	67 156	100
26	REC_NO	N° de registros del paciente	1973	2008	35	67 156	100
27	TYPEFUP	Tipo de seguimiento	1973	2008	35	67 156	100
28	AGE_REC	Grupo de edad del paciente	1973	2008	35	67 156	100
29	AJ_3SEER	Etapa de la enfermedad AJCC	1988	2003	15	35 698	100
30	NUMPRIMS	Número de tumores primarios	1973	2008	35	67 156	100
31	STCOUNTY	Código Estado-Condado	1973	2008	35	67 156	100
32	SURV_TM	Tiempo de supervivencia	1973	2008	35	67 156	100
33	STAT_REC	Estado de vida del paciente	1973	2008	35	67 156	100
34	DTH_CL	Muerte a causa del cáncer	1973	2008	35	67 156	100
35	O_DTH_CL	Muerte por otra enfermedad	1973	2008	35	67 156	100

Resultados y discusión

Una vez establecidas las variables significativas, se realizó el análisis de variabilidad y distribución de las principales variables oncológicas que registran información puntual sobre el paciente y el cáncer de mama. Estas variables fueron: a) origen del paciente, b) edad del paciente al momento del diagnóstico, c) año de diagnóstico del cáncer, d) tipo de la enfermedad, y e) confirmación del diagnóstico del cáncer.

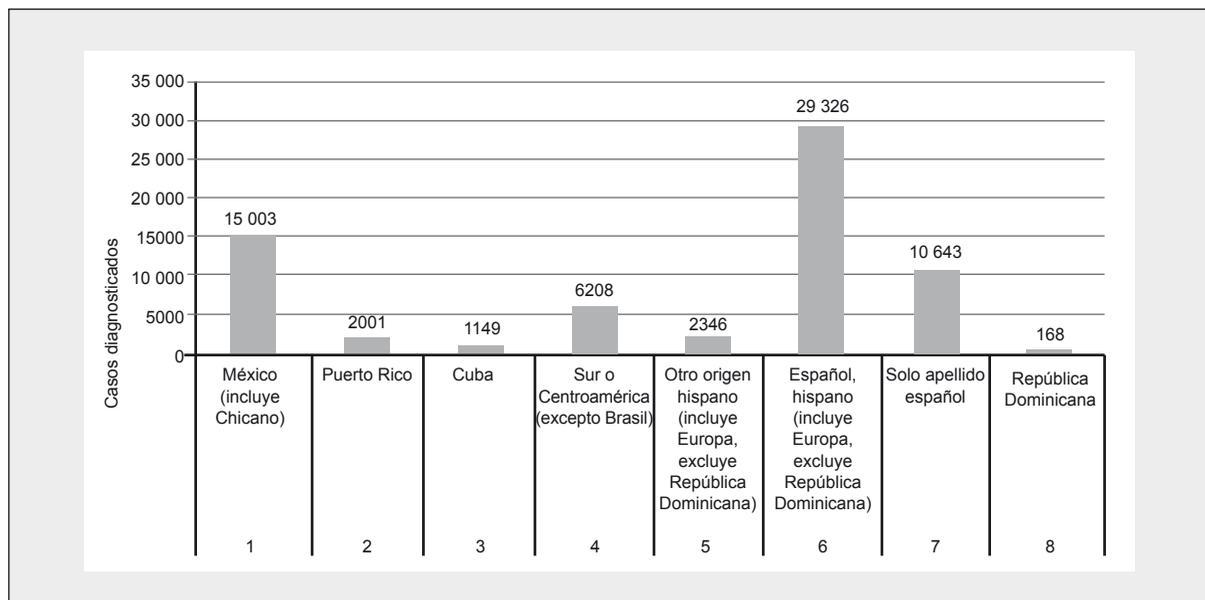


Figura 1. Cáncer de mama de acuerdo con el origen de la paciente.

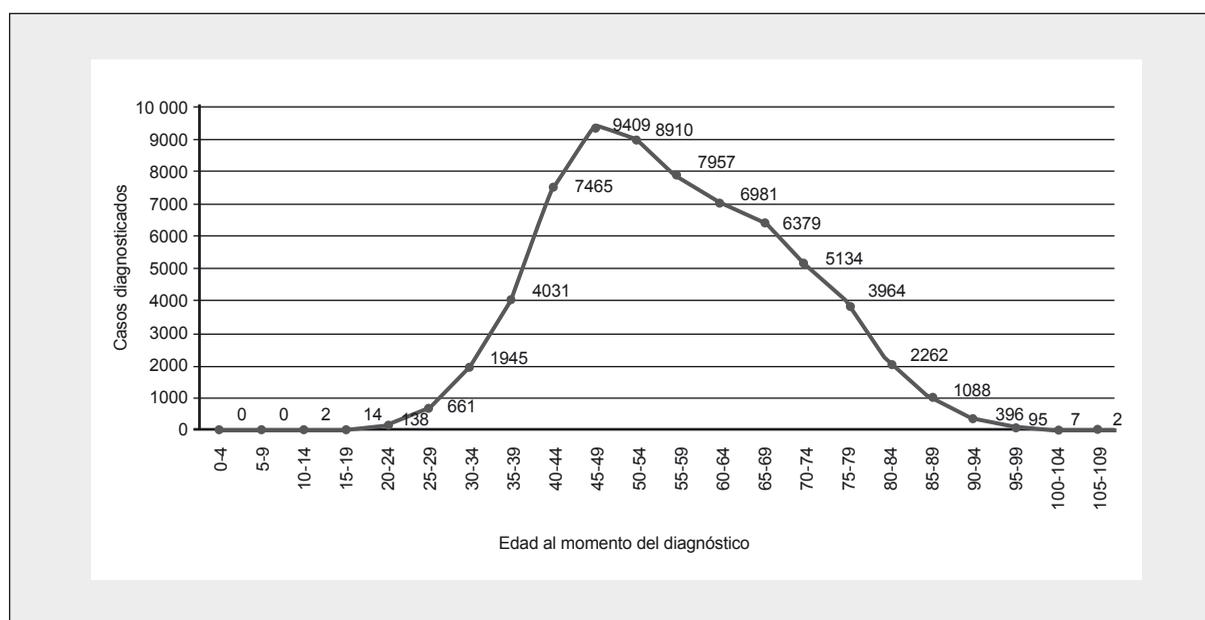


Figura 2. Casos diagnosticados de acuerdo con la edad de la paciente.

Origen de la paciente

Sobre el origen de la paciente (ORIGIN) (Figura 1), se observa que el mayor número de casos de cáncer de mama se presenta en las categorías México (15 003), español, hispano, (29 326) y solo apellido español (10 643). Otro grupo importante de casos diagnosticados se presenta en las categorías Sur o Centroamérica (6208), Puerto Rico (2001), Cuba (1149), otro origen hispano (2346) y República Dominicana (168).

De los casos diagnosticados, se puede inferir que las variadas diferencias del número de casos clínicos registrados, por categoría, son básicamente proporcionales al tamaño de la población de dichas categorías que viven en territorio estadounidense, así como por la fecha de inicio en la que se empezó a registrar la información sobre cada una de las categorías; originando esta variabilidad de una categoría a otra. Esto reafirma el comportamiento no uniforme del cáncer de mama.

Edad de la paciente al momento del diagnóstico

La edad de la paciente en el momento del diagnóstico (AGE_DX) fue otra variable de interés en este estudio (Figura 2). Se observó la presencia de casos de cáncer de mama en mujeres menores de 30 años, llegando incluso a adolescentes de 14 años (dos casos). Otros casos en mujeres menores de 30 años fueron de 15 a 19, con 14; de 20 a 24 con 138, y de 25 a 29, con 661 casos. Esto indica que el cáncer de mama puede aparecer en edades tempranas, en este caso, adolescentes y mujeres jóvenes. Además, hubo un aumento progresivo de los casos diagnosticados con esta enfermedad desde los 30 años, alcanzando el mayor número de casos entre los 45 y 49 años (9409 registros), donde las edades 48 y 49 representan los picos más altos de casos diagnosticados con 1939 y 1936, respectivamente.

A partir de los 50 años también hay una importante presencia de casos diagnosticados, que se hace más evi-

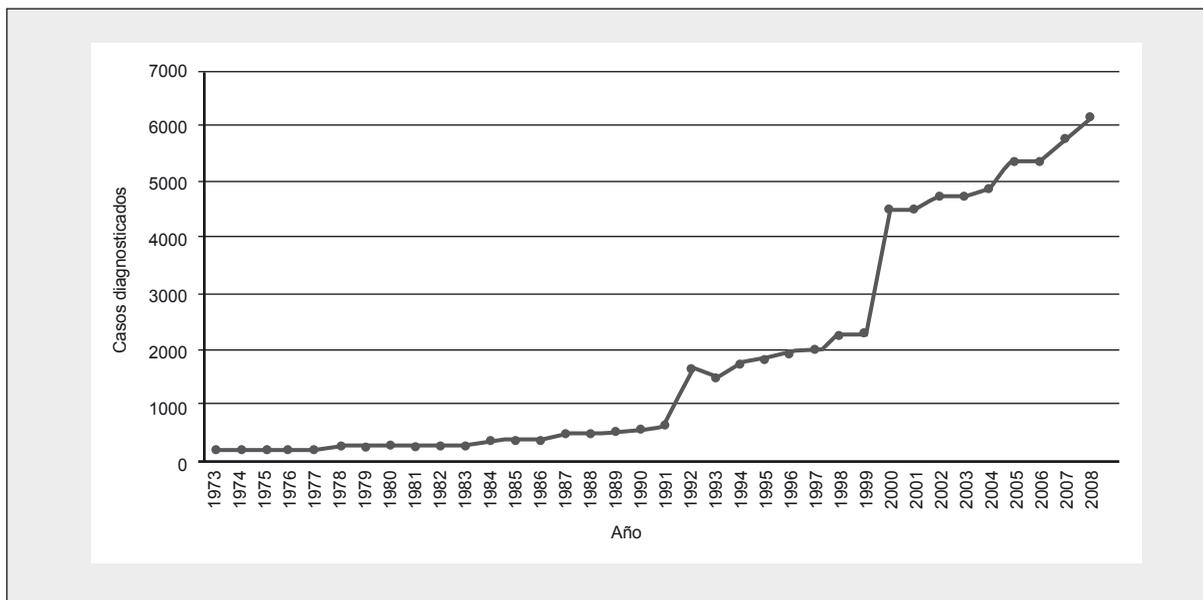


Figura 3. Casos de cáncer de mama de acuerdo con el año de diagnóstico.

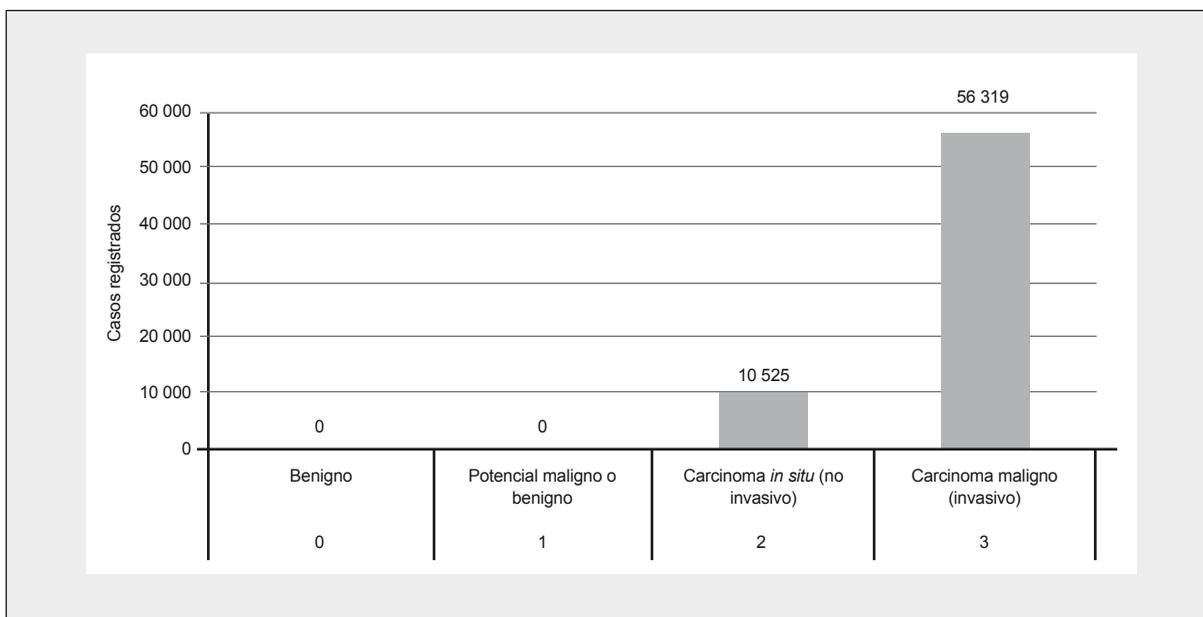


Figura 4. Cáncer de mama por tipo de enfermedad.

dente entre los 50 y los 75 años. A partir de los 75 años, se observa una disminución significativa, que llegan a dos registros de mujeres de entre 105 y 109 años, específicamente, una diagnosticada a los 105 y otra a los 107 años.

Año de diagnóstico del cáncer

Según el año de diagnóstico del cáncer (DATE_YR), se observó que el aumento de la enfermedad fue progresivo de 1973 a 2008 con una tendencia no uniforme (Figura 3), ya que con el tiempo el número de casos detectados ha aumentado significativamente. Esto se debe al incremento de pacientes con esta afección, así como a la expansión del área de cobertura de vigilancia y seguimiento de la enfermedad.

Se observó también que entre 2000 y 2008 el incremento de esta enfermedad fue más notorio, en comparación con años anteriores (1973 a 1999). Es decir, solo en los últimos ocho años (2000-2008) el número de casos

diagnosticados fue 45 950, lo que representa el 68.75% del total de registros disponibles (66 840), mientras que para el período 1973-1999 (26 años) el número total de casos registrados fue de 20 890 (31.25%). Además, se aprecia que, en solo un año, 1999-2000, el número de casos aumentó casi 100%, es decir, de 2280 registros en 1999 a 4490 en 2000. Esta situación puede deberse al aumento de casos y a la expansión del área de cobertura de vigilancia y seguimiento de la enfermedad.

Tipo de enfermedad

La Figura 4 muestra la distribución y variabilidad del tipo de cáncer de mama (BEHO3V), según los tipos: benigno, potencial maligno o benigno, carcinoma *in situ* (no invasivo) y carcinoma maligno (invasivo). En general, se observó que los tipos benigno y potencial maligno o benigno no presentan registros. Además, se verificó que la mayor cantidad de casos registrados se concentra en el tipo carcino-

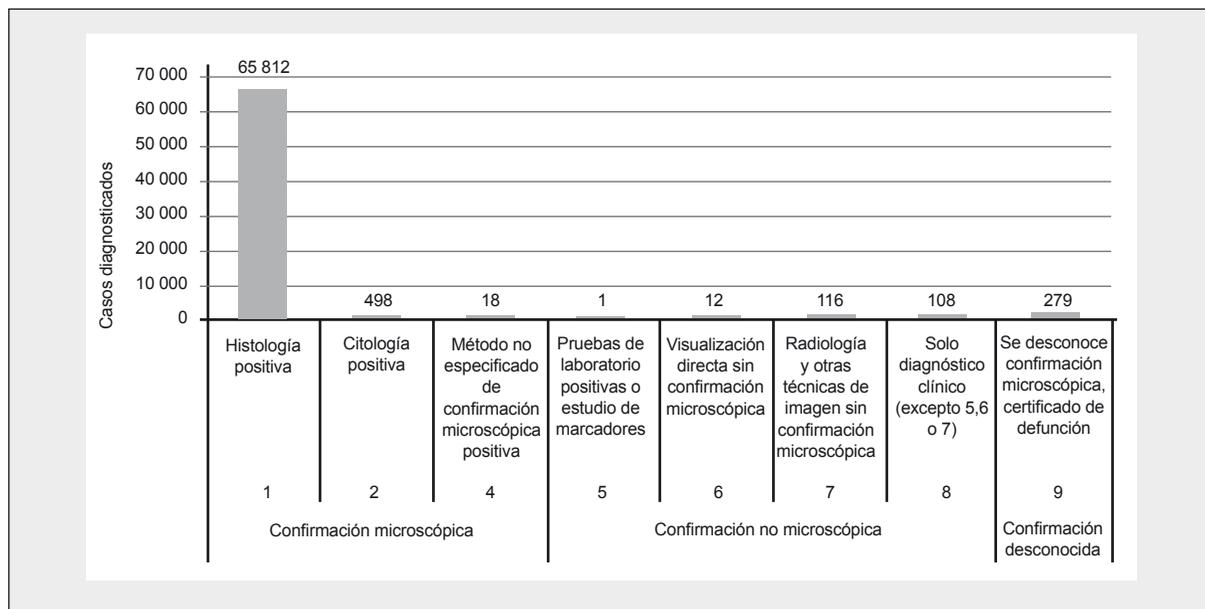


Figura 5. Métodos utilizados para confirmar el diagnóstico de la enfermedad.

ma maligno (invasivo), con 56 319 registros, lo que representa el 84.25% del total de casos disponibles, mientras que los casos restantes (10 521 registros) son del tipo carcinoma *in situ* (no invasivo), lo que representa el 15.75%.

El elevado número de casos de carcinoma invasivo puede deberse a la falta de un diagnóstico oportuno que permita atacar la enfermedad en su etapa inicial (carcinoma no invasivo). Por lo tanto, esta variable es importante para el seguimiento y análisis del comportamiento de la neoplasia, lo que permite a los especialistas definir mejores opciones de tratamiento y recuperación de la paciente.

Confirmación del diagnóstico del cáncer

La Figura 5 muestra la distribución y variabilidad del mejor método utilizado para confirmar la presencia de cáncer de mama (DX_CONF). Los métodos utilizados son: confirmación microscópica (histología positiva, citología positiva, método no especificado de confirmación microscópica positiva); confirmación no microscópica (pruebas de laboratorio positiva, visualización directa sin confirmación microscópica, radiología y otras técnicas de imagen sin confirmación microscópica, solo diagnóstico clínico) y confirmación desconocida (se desconoce confirmación microscópica, certificado de defunción).

Se observó que el mayor número de casos diagnosticados fue mediante confirmación microscópica, principalmente por exámenes de histología, lo que dio un total de 65 812 casos, que representan el 98.46% de los casos diagnosticados. Otros casos, en menor medida, también se confirmaron mediante citología positiva ($n = 498$) y método no especificado ($n = 18$). Se distingue también que un grupo menor de casos diagnosticados fueron confirmados por medio de evaluaciones no microscópicas, como radiología y otras técnicas de imágenes ($n = 116$), diagnóstico clínico ($n = 108$), visualización directa ($n = 12$) y pruebas de laboratorio ($n = 1$). Además, se observó la presencia de 279 casos cuya forma de confirmación es desconocida.

Conclusiones

El análisis de datos es una de las actividades fundamentales en diversos campos de conocimiento de la actividad humana. Los datos analizados fueron registros clínicos de cáncer de mama del programa SEER del NCI de los Estados Unidos.

El análisis de los datos se realizó en dos etapas. En la primera, se llevó a cabo una evaluación preliminar de la disponibilidad de datos de todas las variables incluidas en la base de datos SEER. Esto con el fin de establecer aquellas relevantes según el período de sus registros, descartando las variables con una alta cantidad de valores nulos. En la segunda etapa se determinó la calidad de la serie de datos para establecer las variables significativas asociadas con el cáncer de mama en mujeres de origen hispano.

De la evaluación se seleccionaron 35 variables consideradas significativas, lo que representa el 28.2% del total de variables registradas en la base de datos SEER (124 variables) y el 47.9% del total de variables que fueron elegidas en el análisis preliminar (73 variables).

Luego de la determinación de las variables significativas, se analizó la variabilidad y distribución de las principales variables oncológicas que registran información puntual sobre la paciente y el cáncer de mama, como origen de la paciente, edad, año de diagnóstico, tipo de enfermedad y confirmación del diagnóstico. Este análisis proporcionó la identificación de tendencias y comportamientos de los datos disponibles sobre el cáncer de mama en mujeres de origen hispano.

Por último, se observó que el cáncer de mama es una enfermedad con un comportamiento no uniforme, lo que indica que la posibilidad de recuperación depende de la etapa del cáncer, si se encuentra exclusivamente en la mama o se ha diseminado a otras partes del cuerpo, el tipo de cáncer, el tratamiento recibido, y la salud de la paciente en general. Por lo tanto, un diagnóstico en una etapa temprana de esta enfermedad es crucial.

Bibliografía

1. Keogh E, Lin J. Clustering of time series subsequences is meaningless: implications for previous and future research. *Knowledge and Information Systems* 8(2):154-177, 2005.
2. Kessler M. Apuntes de métodos estadísticos de la ingeniería. Report, Polytechnic University of Cartagena, España; 2005.
3. Puerto J, Paz M. Análisis descriptivo de series temporales aplicadas al precio medio de la vivienda en España. Report, Management Mathematics for European Schools, España; 2001.
4. Chiu B, Keogh E, Lonardi S. Probabilistic discovery of time series motifs. En: ACM SIGKDD international conference on knowledge discovery and data mining, Washington, USA, August 24-27; 2003. Pp. 493-498.
5. Molero G, Céspedes Y, Meda M. Caracterización y análisis de la base de datos de cáncer de mama SEER-DB. En: Ninth International Congress on Informatics in Health, Havana, Cuba, March 18-22; 2013. Pp. 95-103.
6. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Disponible en: <https://seer.cancer.gov/data>. Consultado el 31 de octubre de 2016.
7. National Cancer Institute. About the SEER Program. Disponible en: <http://seer.cancer.gov/about>. Consultado el 2 de noviembre de 2016.
8. National Cancer Institute. Home SEER. Surveillance, Epidemiology and End Results Program. Disponible en: <https://seer.cancer.gov/data/documentation.html>. Consultado el 4 de septiembre de 2016).
9. Centers for Disease Control and Prevention. National Program of Cancer Registries. Disponible en: www.cdc.gov/cancer/npcr/about.htm. Consultado el 18 de septiembre de 2016.
10. American Cancer Society. Cancer Surveillance Programs in the United States. Disponible en: www.cancer.org/cancer/cancer-basics/cancer-surveillance-programs-and-registries-in-the-united-states.html. Consultado el 10 de octubre de 2016.

Información relevante

Tendencias y comportamientos de los datos disponibles sobre el cáncer de mama

Respecto al autor

Guillermo Molero Castillo. Doctor en Tecnologías de Información, Universidad de Guadalajara, México. Maestro en Ciencias de la Computación, Universidad Nacional Autónoma de México, México. Actualmente trabaja como investigador en el Consejo Nacional de Ciencia y Tecnología (CONACYT), comisionado a la Universidad Veracruzana, México. Principales líneas de investigación: ciencia de datos, inteligencia artificial, inteligencia de negocios, interacción humano-computadora y gestión de procesos. Ha trabajado en instituciones públicas y privadas. Ha participado en eventos nacionales e internacionales; es autor de varias publicaciones científicas y actualmente es miembro del Sistema Nacional de Investigadores.



Respecto al artículo

El cáncer de mama es una enfermedad con un comportamiento no uniforme, lo que indica que la posibilidad de recuperación depende de la etapa de la enfermedad, de si se encuentra exclusivamente en la mama o se ha diseminado a otras partes del cuerpo, así como del tipo, el tratamiento recibido y la salud de la paciente en general.

El autor pregunta

El análisis de datos es una actividad fundamental a través de la cual se establece el contacto directo con el problema. El análisis se realizó en dos etapas. La primera consistió en un análisis preliminar del total de variables listadas en la base de datos SEER. Los datos analizados fueron de 1973 a 2008. En la segunda etapa, se determinó la calidad de la serie de datos para establecer las variables significativas asociadas al diagnóstico de cáncer de mama en mujeres de origen hispano.

¿Cuántas variables fueron identificadas como significativas para el diagnóstico de cáncer de mama en el registro de datos SEER?

- A** Treinta variables.
- B** Cuarenta variables.
- C** Treinta y cinco variables.
- D** Veinticinco variables.
- E** Ninguna de las opciones es correcta.

Corrobore su respuesta: www.siicsalud.com/dato/evaluaciones.php/155526

Palabras clave

análisis de datos, estudios de series temporales, neoplasias de la mama, salud de la mujer

Key words

data analysis, time series studies, breast neoplasms, women's health

Lista de abreviaturas y siglas

SEER, programa de Vigilancia, Epidemiología y Resultados Finales; NCI, Instituto Nacional del Cáncer.

Cómo citar

Molero Castillo G, Céspedes González Y, Álvarez Domínguez TP. Tendencias y comportamientos de los datos disponibles sobre el cáncer de mama. *Salud i Ciencia* 23(1):43-50, May-Jun 2018.

How to cite

Molero Castillo G, Céspedes González Y, Álvarez Domínguez TP. Trends and behaviors in available data on breast cancer. Salud i Ciencia 23(1):43-50, May-Jun 2018.

Orientación

Clínica, Diagnóstico

Conexiones temáticas

Los informes de *Salud(i)Ciencia* se conectan de manera estricta (**i**) o amplia (**▶**) con diversas especialidades.

