



UTILIDAD DEL ANÁLISIS ESTADÍSTICO IMPLICATIVO, PARA LA IDENTIFICACIÓN DE FACTORES DE RIESGO DEL CÁNCER DE PRÓSTATA.

Autores: Dr. Yohani García Mederos¹, Dra. Yuannis Peña Almenares².

¹ Esp. 1er grado en Bioestadística y MGI, Dpto. Provincial de Genética Médica. Profesor Asistente e Investigador Agregado. Hospital General Provincial Docente "Carlos Manuel de Céspedes y del Castillo" Granma, Cuba.

² Esp. 1er grado en Oftalmología y MGI. Servicio de Oftalmología, Dpto. Úvea. Profesora Asistente e Investigadora Agregada. Hospital General Provincial Docente "Carlos Manuel de Céspedes y del Castillo" Granma, Cuba.

E-mail primer autor: yohanigarcia78@gmail.com

RESUMEN

Introducción: Desde hace años viene incrementándose el número de hombres con cáncer de próstata en nuestro país, y de forma particular en la provincia de Santiago de Cuba. Esta situación, unida al desarrollo del Análisis Estadístico Implicativo como herramienta para el estudio de la causalidad, nos motivó a la realización de este trabajo, con **el objetivo** de evaluar la utilidad de este nuevo instrumento en la identificación de los factores de riesgo para esta enfermedad. **Materiales y métodos:** Se realizó un estudio analítico de casos y controles en pacientes con neoplasia de próstata, atendidos en los hospitales "Conrado Benítez" y "Dr. Juan Bruno Zayas" de Santiago de Cuba, desde el 1 de julio del 2012 hasta el 31 de junio del 2013. Se seleccionaron 50 casos por muestreo sistemático aleatorio y 50 controles mediante muestreo por conglomerados en 4 etapas, relación (1:1). Ambas técnicas fueron comparadas utilizando como estándar de oro la regresión logística. **Resultados y discusión:** Se obtuvieron elevados valores de sensibilidad, especificidad, valores predictivos, razones de verosimilitud e índices de validez, Youden y Kappa entre ellas. **Conclusiones:** Las dos técnicas se complementan en la identificación de factores de riesgo, permitiendo un mejor estudio e interpretación del problema de la causalidad.

Palabras clave: análisis estadístico implicativo, cáncer de próstata, factor de riesgo, regresión logística, causalidad, sensibilidad, especificidad.



INTRODUCCIÓN

Desde hace algunos años se ha venido evidenciando un aumento en la incidencia del cáncer de próstata como causa de muerte en los hombres de la población cubana en general y particularmente en los de la provincia de Santiago de Cuba. De esta situación se deduce la importancia de frenar esta tendencia, o incluso, revertirla en la medida de lo posible. Para conseguir esto es imprescindible tener bien identificadas las causas o factores que conllevan a su aparición, y trabajar en su prevención desde el nivel primario de atención de la salud.

Varios autores¹⁻³ coinciden en que, a mayor cantidad de factores de riesgo para un cáncer presentes en una misma persona, mayor riesgo tendrá esta de padecer dicha neoplasia; sin embargo, tan solo presentarlos no implica que obligatoriamente el paciente va a desarrollar la enfermedad, debido a esta situación es que las técnicas estadísticas de implicación directa y análisis simétrico de datos, como la correlación y X^2 , no son adecuadas para estudiar este fenómeno. Tradicionalmente, y gracias al desarrollo de la informática y los paquetes estadísticos, las técnicas más utilizadas en estos casos son las multivariadas, siendo las regresiones logísticas (RL) las más esgrimidas.⁴⁻⁵

En la década de los 70's del pasado siglo, el profesor francés Régis Gras del Instituto de Matemática de Nantes, especializado en el estudio de la causalidad y la cuasi-implicación, desarrolló el análisis estadístico implicativo (ASI, por sus siglas en inglés), que se caracteriza por la investigación del fenómeno de la creación de reglas de inferencia imprecisas o parciales, entre variables binarias que describen a una población de individuos, o sea, trata de descubrir reglas inductivas no simétricas para modelar relaciones del tipo "si a entonces casi b".^{6,7}

La seria situación de salud marcada por la elevada incidencia del cáncer de próstata y gracias a la posibilidad que brindan las técnicas del ASI para la determinación de relaciones de causalidad entre variables, se decidió realizar esta investigación.^{8,9}

OBJETIVO

Evaluar la utilidad del análisis estadístico implicativo en la identificación de factores de riesgo para el cáncer de próstata en pacientes santiagueros aquejados de esta enfermedad.

MATERIALES Y MÉTODOS

Se realizó un estudio analítico de casos y controles en pacientes con cáncer de próstata, atendidos en las consultas externas de los hospitales "Conrado Benítez" y "Dr. Juan Bruno Zayas Alfonso" de Santiago de Cuba, desde el 1 de julio del 2012



hasta el 31 de junio del 2013. Se estudiaron además residentes de esta provincia sin la citada afección y de entornos similares a los enfermos.

Fueron seleccionados 50 casos mediante un muestreo sistemático aleatorio con arranque 3, e intervalo de salto 4, para lo cual se tomó un paciente de cada 5 de los que asistieron a la consulta, e igual cantidad de controles según muestreo por conglomerados en 4 etapas, donde las unidades seleccionadas en cada periodo fueron: municipios, áreas de salud, consultorios del médico de la familia y casas, respectivamente. Posteriormente, de cada casa seleccionada se escogió mediante un muestreo simple aleatorio, una persona que cumplía con los criterios requeridos para ser control. Si la vivienda seleccionada no poseía personas elegibles, se tomó de la siguiente casa seleccionada 2 controles hasta completar el tamaño de muestra necesario.

Se analizaron 13 variables, una de respuesta: cáncer de próstata (Cáncer) y 12 explicativas: edad, antecedente genético (PAG), antecedente patológico personal (PAP), tipo corporal obeso (FEEn), altos niveles de testosterona (FEEx), consumo de alcohol (Bebe), consumo de cigarro y/o tabaco (Fuma), vida sedentaria (PVS), dieta inadecuada (PDI), radiaciones ionizantes (PRI), raza y nivel socioeconómico (PS1-3). Las 10 primeras respondían a variables dicotómicas y se codificaron en 0 y 1, las 2 últimas son modales con 3 categorías, para cada una se definieron 3 nuevas variables dicotómicas: raza blanca (PRB), raza mestiza (PRM), raza negra (PRN), nivel socioeconómico bajo (PS1), nivel socioeconómico medio (PS2) y nivel socioeconómico alto (PS3).

La variable edad se dividió en 2 grupos: 50 años o menos (0) y mayores de 50 años (1).

Para la recogida de la información se confeccionó un formulario donde se plasmaron los datos de cada paciente, concernientes a todas las variables del estudio.

Se elaboró una base de datos en MS Excel y para el procesamiento y presentación de los resultados se emplearon los sistemas EPIDAT 3.1, SPSS versión 20 y CHIC versión 5.0.

Primeramente, se utilizó la regresión logística binaria para las variables no correlacionadas, identificándose los factores de riesgo que tuvieron significación estadística. Finalmente se empleó el método por pasos hacia adelante ya que fue el que mejor ajuste al modelo presentó, según la prueba de Hosmer y Lemeshow. A continuación, mediante el ASI se identificaron las variables que tuvieron relación causal significativa y que, por tanto, constituían factores de riesgo para contraer la enfermedad.



Posteriormente, para la comparación de las dos técnicas, fueron abordadas como “medios diagnósticos” para la identificación de factores de riesgo, se empleó como estándar de oro la regresión logística binaria, considerándose cada variable como factor de riesgo o no según los resultados de las 2 técnicas, conformando la tabla de 2x2 correspondiente; se calcularon la sensibilidad, especificidad, valores predictivos, razones de verosimilitud, así como los índices de validez, Youden y Kappa entre ellas.

También se identificaron las similitudes y diferencias entre estas técnicas.

Todos los datos incluidos en el estudio fueron debidamente aprobados por el jefe provincial del programa y aquellos relacionados con los pacientes se obtuvieron previa autorización de estos.

RESULTADOS

En la Tabla 1 se observa el predominio, en los casos, de las variables antecedentes genéticos, raza mestiza, dieta inadecuada, nivel socioeconómico alto y antecedentes personales.

Tabla 1. Casos y controles según factor de riesgo

Factor de riesgo	Casos		Controles	
	No	%	No	%
Edad (EDAD)	49	55,1	40	44,9
Raza blanca (PRB)	18	46,2	21	53,8
Raza mestiza (PRM)	23	74,2	8	25,8
Raza negra (PRN)	9	30,0	21	70,0
Antecedentes genéticos (PAG)	31	91,2	3	8,8
Antecedentes patológicos personales (PAP)	12	63,2	7	36,8
Factores endocrinos endógenos (PEEn)	8	50,0	8	50,0
Factores endocrinos exógenos (PEEx)	1	100,0	0	0,0
Consumo de bebidas alcohólicas (BEBE)	27	56,2	21	43,8
Hábito de fumar (FUMA)	32	47,8	35	52,2
Vida sedentaria (PVS)	38	53,5	33	46,5
Dieta inadecuada (PDI)	21	70,0	9	30,0
Radiación ionizante (PRI)	2	66,7	1	33,3
Nivel socioeconómico bajo (PS1)	5	45,4	6	54,6
Nivel socioeconómico medio (PS2)	34	46,6	39	53,4
Nivel socioeconómico alto (PS3)	11	68,7	5	31,3



Resultados de la RL

Para la muestra estudiada los factores de riesgo según la RL binaria, con un nivel de confianza del 95% fueron: la edad (EDAD), ya que ser mayor de 50 años incrementó en 43,73 veces la probabilidad de tener esta enfermedad; el antecedente genético (PAG) que aumentó en 37,22 veces la probabilidad de padecer este cáncer en aquellos que lo presentaron y la raza mestiza (PRM) con un odd ratio de 7,28 por lo que las personas de este color tienen 7,28 veces más riesgo de desarrollar esta neoplasia que las de otra raza (Tabla 2).

Tabla 2. Resultados de la regresión logística binaria

Variables	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
EDAD	3,78	1,48	6,52	1	0,011	43,73	2,41	794,84
PRM	1,98	0,64	9,64	1	0,002	7,28	2,08	25,46
PAG	3,62	0,81	19,98	1	0,000	37,22	7,62	181,79
Constante	-5,05	1,54	10,74	1	0,001	0,01		

Resultados del ASI

El Árbol de similaridad de la Figura 1 muestra la agrupación de los factores de riesgo según el índice de similaridad de Lerman y los nodos significativos (en trazos más gruesos de color rojo), que se forman en los niveles 1, 5, 8, 12 y 14, siendo el nivel 1 el más significativo. Se aprecia la formación de 2 grandes clases, en una de las cuales no está contenida la variable dependiente, en este caso Cáncer de próstata, por tanto, las variables en dicha clase no tienen semejanza o correspondencia directa con la variable dependiente, estas son, Raza blanca y negra, Exposición a radiaciones ionizantes, Antecedente patológico personal, Factores endocrinos endógenos y exógenos y Nivel socioeconómico alto.

En la otra clase se forman 4 subgrupos:

- Un primer subgrupo formado por la unión de las variables cáncer (CÁNCER) y antecedentes genéticos (PAG) con un índice de similaridad de 100% y la formación de un nodo significativo de nivel 1 de la jerarquía, que a su vez se unen formando un nodo significativo nivel 5 y un índice de similaridad de 81% con la variable raza mestiza (PRM). Este subgrupo muestra que generalmente los pacientes con este cáncer tienen antecedentes genéticos de neoplasia y son de raza mestiza.



- Un segundo subgrupo formado por la unión de las variables dieta inadecuada (PDI) y nivel socioeconómico bajo (PS1), con un índice del 98%, que se unen al primer subgrupo formando un nodo significativo de nivel 8. Este subgrupo muestra que los individuos con dieta inadecuada habitualmente pertenecen al nivel socioeconómico bajo.
- El tercer subgrupo se formó por la unión de las variables vida sedentaria (PVS) y nivel socioeconómico medio (PS2) con una similaridad del 45%. Este subgrupo se vinculó a la unión de los subgrupos anteriores. Aquí se muestra que los pacientes con vida sedentaria tienen un nivel socioeconómico medio.
- El cuarto subgrupo está formado por la unión de las variables bebe (BEBE) y fuma (FUMA) con un índice de similaridad de 69%, que a su vez se unen con la variable edad (EDAD) con un 46% de similaridad, formándose un nodo significativo de nivel 12. O sea, en la muestra estudiada las personas mayores de 50 años fueron en su mayoría bebedores habituales y fumadores.

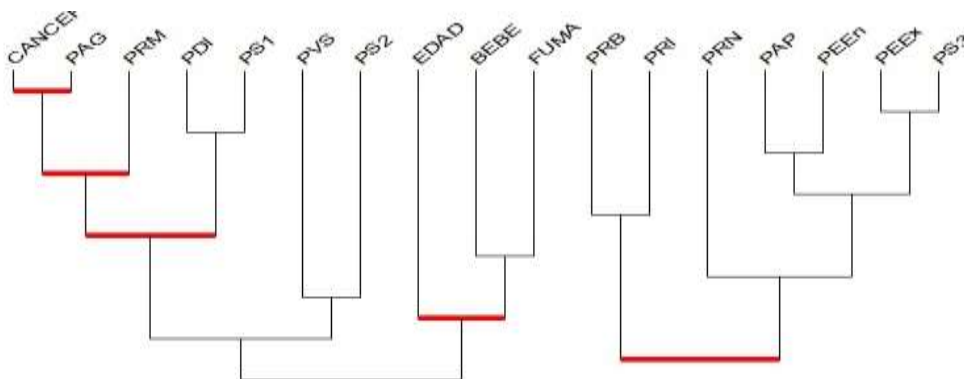


Fig. 1. Agrupación de factores de riesgo según índice de similaridad (Árbol de similaridad).

En la Figura 2 aparece el árbol cohesitivo donde se observan 6 grupos de variables que conforman nodos significativos en los niveles 1, 4, y 6 de jerarquía, siendo el nivel 1 el más significativo. Hay una variable que queda sin agrupar. El análisis de estos grupos en el orden en que aparecen indica que:

- Pertenecer a la raza mestiza (PRM) implicó con un índice de cohesión del 82%, que el individuo fuma (FUMA) y la presencia de estas 2 características implicó con una cohesión del 40% que esa persona es un bebedor habitual (BEBE), esta implicación generó un nodo significativo de nivel 6.
- Tener antecedentes genéticos (PAG) implicó, con una cohesión del 100% (la más fuerte de la muestra), que el paciente tenía cáncer (CÁNCER); ahora, tener antecedentes genéticos y cáncer, con un índice de cohesión del 80%



- implica que el individuo es mayor de 50 años (EDAD). Ambas implicaciones generaron respectivamente nodos significativos de nivel 1 y 4 de la jerarquía.
- Ser obeso (PEEn), implicó la presencia de antecedentes patológicos personales (PAP) con un 57% de cohesión, a su vez, la unión de estas 2 características implicó con un índice del 38% que la persona es de raza blanca (PRB).
 - Presentar altos niveles de testosterona (PEEx) implicó que la persona tiene un nivel socioeconómico alto (PS3), con una cohesión de 16%.
 - La presencia de una dieta inadecuada (PDI) trajo consigo, con una cohesión del 94%, la presencia de vida sedentaria (PVS), la unión de estas 2 variables implicó con un índice del 23% la pertenencia al nivel socioeconómico bajo (PS1).
 - La variable radiación ionizante (PRI) queda sin agrupar, esto significa que en la muestra estudiada esa es la variable que menos tiene que ver con el resto.
 - La pertenencia al nivel socioeconómico medio (PS2) implicó con un índice de cohesión de 15% que la persona es de raza negra (PRN).

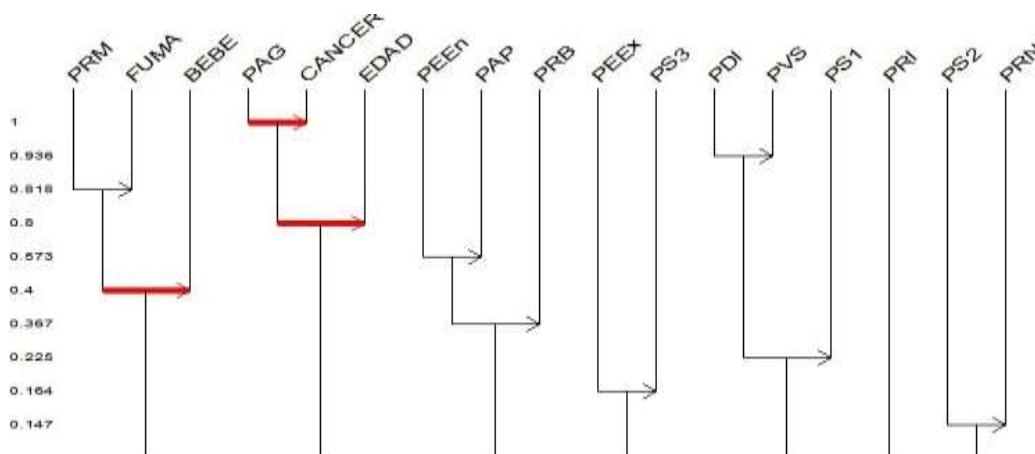


Fig. 2. Reglas y meta reglas que indican relaciones causales entre factores de riesgo (árbol cohesitivo).

En el grafo implicativo de la Fig. 3 se observa la formación de varios caminos:

C1: PS1 → PDI → CANCER, este camino nos muestra que cuando el paciente tiene un nivel socioeconómico bajo va a tener una dieta inadecuada, con una intensidad implicativa comprendida entre el 88 y el 92,99%, y que estos a su vez van a poseer



cáncer de próstata, con una intensidad implicativa entre el 93% y el 95,99%.

C2: PAG → CÁNCER, este camino nos muestra que los pacientes con antecedentes genéticos de primera línea de cáncer de próstata, desarrollaran esta enfermedad con una intensidad de implicación comprendida entre el 99 y el 100%.

C3: PRM → CÁNCER, esta vía muestra que los pacientes de raza mestiza desarrollaron esta enfermedad con una intensidad de implicación comprendida entre el 96 y el 98,99%.

C4: CÁNCER → EDAD, a su vez este camino muestra lo poco probable que es que un paciente que padezca cáncer de próstata no sea mayor de 50 años, con una intensidad de implicación entre el 96 y el 98,99%.

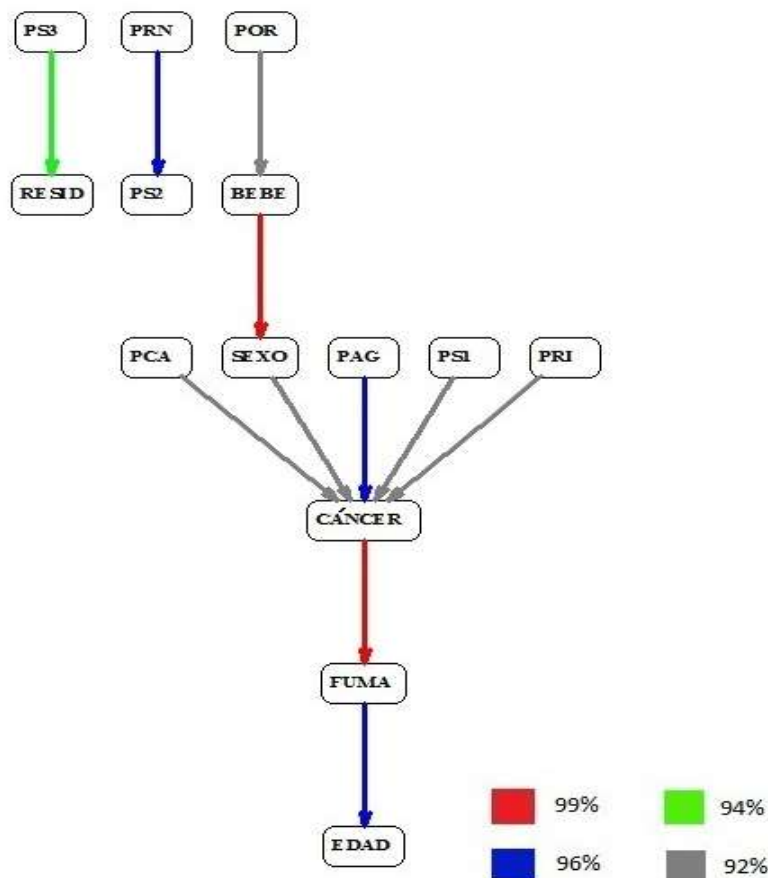


Fig. 3. Grafo implicativo



DISCUSIÓN

Este estudio corrobora por ambas técnicas lo planteado en la literatura, de que los únicos factores reconocidos son la edad, el grupo racial y la historia familiar de la enfermedad⁴, además el ASI detecta 2 factores de riesgo más: la dieta inadecuada y el nivel socioeconómico bajo.

Ambas técnicas clasificaron la edad mayor de 50 años como factor de riesgo con un odd ratio de 43,733, pudiendo incrementar en esta cantidad de veces la probabilidad de padecer cáncer de próstata en aquellos que son mayores de 50 años, que en los que no lo son, resultados similares obtuvieron Brathwaite y Herranz^{10,11}.

El antecedente genético también fue clasificado como un factor de riesgo para este cáncer, en consonancia con otros autores^{4,12,13}, capaz de aumentar el riesgo de aparición de esta enfermedad en 37,219 veces en aquellos individuos que verifiquen esta característica.

La raza es otro de los factores que se invocan en la etiología de esta enfermedad^{11,13}. Este estudio concluyó que para la muestra estudiada los individuos de la raza mestiza tuvieron 7,276 veces más probabilidad de padecer cáncer de próstata que los de otro color de piel. Otros autores como Pérez⁸, recogen que la raza más afectada es la negra, mientras que otros como Fernández⁴ dividen la muestra en sólo 2 grupos raciales, blancos y negros o mestizos. En nuestro país y más en la provincia donde se aplicó el estudio, existe un gran mestizaje, he aquí la posible causa del resultado obtenido.

La dieta inadecuada constituye un factor de riesgo con el que no todos los autores están de acuerdo, no obstante, nosotros concordamos con los que lo plantean como agente causal^{8,13}. En este estudio la RL no lo clasifica como tal, pero el ASI si lo detecta con un 93% de confianza.

Según Herranz¹¹, la enfermedad guarda una relación variable con el nivel socioeconómico, en nuestra investigación, con el empleo del ASI, la relación entre este cáncer y el bajo poder adquisitivo fue significativa. Esto podría deberse a que las personas con bajos ingresos tienen menos oportunidades de tener una dieta balanceada lo que redundo en un aumento del riesgo.

Aún hoy se realizan estudios para definir si factores como la vasectomía, la hiperplasia prostática benigna, la obesidad, el sedentarismo, el tabaquismo, la exposición a radiaciones, la industria del caucho, los altos niveles de testosterona, la exposición al cadmio o los virus de transmisión sexual puedan tener alguna importancia en la génesis de dicha enfermedad⁸. No obstante, ninguna de estas variables resultó significativa en este estudio.

La comparación de los resultados obtenidos por ambas técnicas arrojó que mediante la RL se identificaron como factores de riesgo las variables edad, raza mestiza y



antecedentes genéticos. El ASI identificó estas mismas variables y además la dieta inadecuada y el nivel socioeconómico bajo.

La técnica del ASI en el cáncer de próstata arrojó una capacidad para detectar las variables significativamente asociadas al cáncer ya sea como factor de riesgo o como factor protector, entre un 83,33 y un 100%. Demostró además entre un 61,16 y un 100,00% de capacidad para identificar las variables que no constituyen factor de riesgo.

Con esta técnica es 6,5 veces más probable detectar un factor de riesgo que con la RL. Así mismo, es improbable que una variable que no sea factor de riesgo sea clasificada como agente causal. Los índices obtenidos de validez, o proporción correcta de aciertos, y de Youden, que mide la eficiencia del medio diagnóstico, fueron elevados 87,50% y 0,85 respectivamente y se consideran como buenos, así mismo el índice Kappa de concordancia entre ambas técnicas fue de 0,67 el cual también se considera como bueno.¹⁴

En cuanto al nivel de confianza, en el ASI se puede modificar según el criterio del investigador, de manera que si se baja este nivel aparecen nuevas variables en la relación de implicación, mientras que en la RL se mantienen las mismas variables con el mismo *odd ratio*, solo se hace más estrecho o más amplio el intervalo de confianza para la exponencial de beta. La primera, da las relaciones entre la variable respuesta y las covariables, pero además facilita la correspondencia entre las propias variables independientes, y la segunda, solo proporciona la relación entre la variable respuesta y las independientes.³

En la presentación de los resultados el ASI muestra 3 gráficos que ilustran mejor los índices obtenidos, mientras que la RL carece de un apoyo gráfico a la información que brinda.

Ambas técnicas brindan diferentes tipos de indicadores, puesto que la RL proporciona el *odd ratio* y el ASI determina los índices de similaridad, cohesión e implicación, por lo que se complementan y permiten una interpretación más completa del problema de la causalidad.

CONCLUSIONES

Se concluye que el análisis estadístico implicativo es una técnica adecuada que complementa a la regresión logística en la identificación de factores de riesgo, logrando una mejor comprensión de la causalidad.



BIBLIOGRAFÍA

1. Fernández L, Galán Y, Jiménez R, Gutiérrez A, Guerra M, Pereda CM et al. Estudio de casos y controles sobre factores de riesgo de cáncer de próstata. Rev Cubana Salud Pública v.31 n.3 Ciudad de La Habana jul.-sep. 2005.
2. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo N. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de próstata. MEDISAN [revista en internet] 2018 [citado 21 de enero 2019]; 22(1): 5. Disponible en: <http://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=78550>
3. García Y, Zamora L, Sagaró N. Análisis estadístico implicativo en la identificación de factores de riesgo en pacientes con cáncer de pulmón. MEDISAN [revista en internet] 2015 [citado 21 de enero 2019]; 19(8): 944. Disponible en: <http://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=60754>.
4. Moraga A, Zamora L, Sagaró N, Moraga A, Rodríguez A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de próstata. En: Rev. MEDISAN 2018; 22(1):51. Disponible en: <https://www.researchgate.net/publication/324844193>.
5. Molina E, Barceló C, Bonito LA, Del Puerto C. Factores de riesgo de cáncer pulmonar en Ciudad de La Habana. En: Rev Cubana Hig Epidemiol v.34 n.2; Ciudad de la Habana Mayo-Ago.1996.
6. Gras R, Couturier R. Especificidades da Análise Estatística Implicativa em relação a outras medidas de qualidade de regras de associação. Rev. Educ. Matem. Pesq., São Paulo, v.15, n.2, pp.249-291, 2013. Disponible en: <https://www.researchgate.net/publication/256293689>.
7. Sagaró N, Zamora L. Evolución histórica de las técnicas estadísticas y las metodologías para el estudio de la causalidad en ciencias médicas. En: Rev. MEDISAN 2019;23(3):534. Disponible en: <https://www.researchgate.net/publication/333703615>.
8. Sagaró N, Zamora L. Why Use the Implicative Statistical Analysis in Studies of Health Causality?. Revista Cubana de Informática Médica 2019;11(1)88-103. Disponible en: <https://www.researchgate.net/publication/333934334>.
9. Paez Y, Sagaró N, Zamora L. Análisis estadístico implicativo en la determinación de factores pronósticos del estado nutricional del paciente grave al egreso. En: Rev. MEDISAN 2018; 22(6):650. Disponible en:



<https://www.researchgate.net/publication/325933468>.

10. Brathwaite AF, Brathwaite N, Del Riego A. Epidemiological profile of cancer for Grand Bahama residents. In: West Indian med. j. vol.56 no.1 Mona Jan. 2007. (N. del T.: En español: En:)

11. Herranz F, Arias F, Arrizabalaga M, Calahorra FJ, Carballido J, et al. El cáncer de próstata en la comunidad de Madrid en el año 2000. En: Actas Urol Esp v.27 n.5. Madrid; 2003. <http://dx.doi.org/10.4321/S0210-48062003000500001>.

12. Pérez R, Fajardo M, Orlandi N. El cáncer de próstata: un desafío constante para el Médico de Familia. En: Rev Cubana Med Gen Integr v.18 n.5; Ciudad de La Habana sep.-oct. 2002.

13. Suárez MP, Michelsen J. El papel del selenio y la vitamina E en la prevención y tratamiento del cáncer de próstata. En: Rev. costarric. salud pública v.13 n.24 San José. 2004.

14. Fuente-Alba C. Silva, Molina Villagra M. Likelihood ratio (razón de verosimilitud): definición y aplicación en Radiología. En: Revista Argentina de Radiología, vol. 81, núm. 3, pp. 204-208, 2017. DOI: <https://doi.org/10.1016/j.rard.2016.11.002>.

Los autores certifican la autenticidad de la autoría declarada, así como la originalidad del texto.